The Astrophysical Journal, 736:19 (22pp), 2011 July 20 doi:10.1088/0004-637X/736/1/19

CHARACTERISTICS OF PLANETARY CANDIDATES OBSERVED BY *KEPLER*. II. ANALYSIS OF THE FIRST FOUR MONTHS OF DATA

William J. Borucki[1], David G. Koch[1], Gibor Basri[2], Natalie Batalha[3], Timothy M. Brown[4], Stephen T. Bryson[1], Douglas Caldwell[5], Jørgen Christensen-Dalsgaard[6], William D. Cochran[7], Edna DeVore[5], Edward W. Dunham[8], Thomas N. Gautier III[9], John C. Geary[10], Ronald Gilliland[11], Alan Gould[12], Steve B. Howell[13], Jon M. Jenkins[5], David W. Latham[10], Jack J. Lissauer[1], Geoffrey W. Marcy[2], Jason Rowe[1], Dimitar Sasselov[10], Alan Boss[14], David Charbonneau[10], David Ciardi[15], Laurance Doyle[5], Andrea K. Dupree[10], Eric B. Ford[16], Jonathan Fortney[17], Matthew J. Holman[10], Sara Seager[18], Jason H. Steffen[19], Jill Tarter[5], William F. Welsh[20], Christopher Allen[21], Lars A. Buchhave[10], Jessie L. Christiansen[5], Bruce D. Clarke[5], Santanu Das[22], Jean-Michel Désert[10], Michael Endl[7], Daniel Fabrycky[17], Francois Fressin[10], Michael Haas[1], Elliott Horch[23], Andrew Howard[2], Howard Isaacson[2], Hans Kjeldsen[6], Jeffery Kolodziejczak[24], Craig Kulesa[25], Jie Li[5], Philip W. Lucas[26], Pavel Machalek[5], Donald McCarthy[25], Phillip MacQueen[7], Søren Meibom[10], Thibaut Miquel[27], Andrej Prsa[28], Samuel N. Quinn[10], Elisa V. Quintana[5], Darin Ragozzine[10], William Sherry[13], Avi Shporer[4], Peter Tenenbaum[5], Guillermo Torres[10], Joseph D. Twicken[5], Jeffrey Van Cleve[5], Lucianne Walkowicz[2], Fred C. Witteborn[21], and Martin Still[5]

[1] NASA Ames Research Center, Moffett Field, CA 94035, USA; William.J.Borucki@nasa.gov
[2] Department of Astronomy, University of California, Berkeley, CA 94720, USA
[3] Department of Physics and Astronomy, San Jose State University, San Jose, CA 95192, USA
[4] Las Cumbres Observatory Global Telescope, Goleta, CA 93117, USA
[5] SETI Institute, Mountain View, CA 94043, USA; Martin.Still@nasa.gov
[6] Danish Asteroseismology Centre, Aarhus University, Aarhus, Denmark
[7] McDonald Observatory, University of Texas at Austin, Austin, TX 78712, USA
[8] Lowell Observatory, Flagstaff, AZ 86001, USA
[9] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
[10] Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA
[11] Space Telescope Science Institute, Baltimore, MD 21218, USA
[12] Lawrence Hall of Science, Berkeley, CA 94720, USA
[13] NOAO, Tucson, AZ 85719, USA
[14] Carnegie Institution of Washington, Washington, DC 20015, USA
[15] Exoplanet Science Institute/Caltech, Pasadena, CA 91125, USA
[16] Department of Astronomy, University of Florida, Gainesville, FL 32611, USA
[17] Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064, USA
[18] MIT, Cambridge, MA 02139, USA
[19] Fermilab, Batavia, IL 60510, USA
[20] Department of Astronomy, San Diego State University, San Diego, CA 92182, USA
[21] Orbital Sciences Corp., Mountain View, CA 94043, USA; Fred.C.Witteborn@nasa.gov
[22] University Affiliated Research Center, University of California, Santa Cruz, CA 95064, USA
[23] Southern Connecticut State University, New Haven, CT 06515, USA
[24] MSFC, Huntsville, AL 35805, USA
[25] Department of Astronomy and Steward Observatory, University of Arizona, Tucson, AZ 85721, USA
[26] Centre for Astrophysics Research, Science & Technology Research Institute, University of Hertfordshire, Hatfield, UK
[27] Centre National d'Etudes Spatiales (CNES—French Space Agency), Toulouse, France
[28] Department of Astronomy, Villanova University, Villanova, PA 19085, USA
Received 2011 January 3; accepted 2011 March 10; published 2011 June 29

ABSTRACT

On 2011 February 1 the *Kepler* mission released data for 156,453 stars observed from the beginning of the science observations on 2009 May 2 through September 16. There are 1235 planetary candidates with transit-like signatures detected in this period. These are associated with 997 host stars. Distributions of the characteristics of the planetary candidates are separated into five class sizes: 68 candidates of approximately Earth-size ($R_p < 1.25\,R_\oplus$), 288 super-Earth-size ($1.25\,R_\oplus \leqslant R_p < 2\,R_\oplus$), 662 Neptune-size ($2\,R_\oplus \leqslant R_p < 6\,R_\oplus$), 165 Jupiter-size ($6\,R_\oplus \leqslant R_p < 15\,R_\oplus$), and 19 up to twice the size of Jupiter ($15\,R_\oplus \leqslant R_p < 22\,R_\oplus$). In the temperature range appropriate for the habitable zone, 54 candidates are found with sizes ranging from Earth-size to larger than that of Jupiter. Six are less than twice the size of the Earth. Over 74% of the planetary candidates are smaller than Neptune. The observed number versus size distribution of planetary candidates increases to a peak at two to three times the Earth-size and then declines inversely proportional to the area of the candidate. Our current best estimates of the intrinsic frequencies of planetary candidates, after correcting for geometric and sensitivity biases, are 5% for Earth-size candidates, 8% for super-Earth-size candidates, 18% for Neptune-size candidates, 2% for Jupiter-size candidates, and 0.1% for very large candidates; a total of 0.34 candidates per star. Multi-candidate, transiting systems are frequent; 17% of the host stars have multi-candidate systems, and 34% of all the candidates are part of multi-candidate systems.

Key words: planetary systems – stars: statistics – planets and satellites: detection – surveys

Online-only material: machine-readable tables

1. INTRODUCTION

Kepler is a discovery-class mission designed to determine the frequency of Earth-size planets in and near the habitable zone (HZ) of solar-type stars. Details of the *Kepler* mission and instrument can be found in Koch et al. (2010a), Jenkins et al. (2010c), and Caldwell et al. (2010). All data through 2009 September 16 are now available through the Multi-Mission Archive (MAST[29]) at the Space Telescope Science Institute for analysis by the community.

Based on the first 43 days of data, five exoplanets with sizes between 0.37 and 1.6 Jupiter radii and orbital periods from 3.2 to 4.9 days were recognized and then confirmed by radial velocity (RV) observations during the 2009 observing season (Borucki et al. 2010; Koch et al. 2010b; Dunham et al. 2010; Jenkins et al. 2010a; Latham et al. 2010). Ten more planets orbiting a total of three stars have subsequently been announced (Holman et al. 2010; Torres et al. 2011; Batalha et al. 2011; Lissauer et al. 2011a).

Because of great improvements to the data-processing pipeline, many more candidates are visible than in the data considered in the papers published in early 2010. When *Kepler*'s first major exoplanet data release occurred on 2010 June 15, 706 target stars had candidate exoplanets (Borucki et al. 2011). In this data release we identify 997 stars with a total of 1235 planetary candidates that show transit-like signatures in the first 132 days of data. A list of false positive events found in the released data is also included in Table 4 with a brief note explaining the reason for classification as a false positive. All false positives are also archived at the MAST. A total of 1202 planetary candidates are discussed herein.

The algorithm that searches for patterns of planetary transits also finds stars with multiple planet candidates. A separate paper presents an analysis of five of these candidates (Steffen et al. 2010). Data and search techniques capable of finding planetary transits are also very sensitive to eclipsing binary (EB) stars, and indeed the number of EBs discovered with *Kepler* exceeds the number of planetary candidates. With more study, some of the current planetary candidates might also be shown to be EBs and some planetary candidates or planets might be discovered orbiting some of the EBs. Prsa et al. (2011) present a list of EBs with their basic system parameters that have been detected in these early data.

2. DESCRIPTION OF THE DATA

Data for all stars are recorded at a cadence of one per 29.4244 minutes (hereafter long cadence or LC). Data for a subset of up to 512 stars are also recorded at a cadence of one per 58.85 s (hereafter short cadence or SC), sufficient to conduct asteroseismic observations needed for the measurements of the stars' sizes, masses, and ages. The results presented here are based only on LC data. For a full discussion of the LC data and their reduction, see Jenkins et al. (2010b, 2010c). See Gilliland et al. (2010) for a discussion of the SC data.

The results discussed in this paper are based on three data segments: the first segment (labeled Q0) started on JD 2454953.53 and ended on 2454963.25 and was taken during commissioning operations, the second data segment (labeled Q1) taken at the beginning of science operations that started on JD 2454964.50 and finished on JD 2454997.99, and a third segment (labeled Q2) starting on JD 2455002.51 and finishing on JD 2455091.48. The durations of the segments are 9.7, 33.5, and 89.0 days, respectively. The observations span a total period of 137.95 days including the gaps. A total of 156,097 LC targets in Q1, and 166,247 LC and 1492 SC targets in Q2 were observed. The stars observed in Q2 were mainly a superset of those observed in Q1. These data have been processed with Science Operations Center pipeline version 6.2 and archived at the MAST. Originally, the bulk of these data were scheduled for release on 2011 June 15, but the exoplanet targets are being released early, so 165,470 LC and 1478 SC targets became available to the public on 2011 February 1. The remaining few targets have a proprietary user other than the *Kepler* science team (e.g., guest observers). Data for these targets will become public by 2011 June 15. The current release date and the proprietary owner for each target are posted at MAST as soon as the data enter the archive, which occurs about four months after data acquisition for the quarter in question is complete.

The results reported here are for the LC observations of 153,196 stars observed during Q2. Other stars were giants or super-giants, did not have valid parameter values, or were in some way inappropriate to the discussion of the exoplanet search. The enlarged set of stars observed in Q2 included most of the stars observed in Q1 and additional stars due to the more efficient use of the available pixels. The selected stars are primarily main-sequence dwarfs chosen from the *Kepler* Input Catalog[30] (KIC). Targets were chosen to maximize the number that were both bright and small enough to show detectable transit signals for small planets in and near the HZ (Gould et al. 2003; Batalha et al. 2010a). Most stars were in the *Kepler* magnitude range $9 < K_p < 16$. The *Kepler* passband covers both the *V* and *R* photometric passbands (Figure 1 in Koch et al. 2010a). See the discussion in Batalha et al. (2010b).

2.1. Noise Sources in the Data

The *Kepler* photometric data contain a wide variety of both random and systematic noise sources. These sources and others are discussed in Jenkins et al. (2010b) and Caldwell et al. (2010). Work is underway to improve the mitigation and flagging of the affected data. Stellar variability over the periods similar to transit durations is also a major source of noise.

Because of the complexity of the various small effects that are important to the quality of the *Kepler* data, prospective users of *Kepler* data are strongly urged to study the data release notes (available at the MAST) for the data sets they intend to use. Note that the *Kepler* data analysis pipeline was designed to perform differential photometry to detect planetary transits, so other uses of the data products require caution.

2.2. Distinguishing Planetary Candidates from False Positive Events

The search for planets starts with a search of the time series of each star for a pattern that exceeds a detection threshold commensurate with a non-random event. Observed patterns of transits consistent with those from a planet transiting its host star are labeled "planetary candidates." (In a few cases, a single drop in brightness that had a high signal-to-noise ratio (S/N) and was of the form of a transit was sufficient to identify a planetary candidate.) Those that were at one time considered to be planetary candidates but subsequently failed some consistency test are labeled "false positives." After passing

[29] http://archive.stsci.edu/Kepler/data_search/search.php

[30] http://archive.stsci.edu/Kepler/Kepler_fov/search.php

all consistency tests described below, and only after a review of all the evidence by the entire *Kepler* Science Team, does the candidate become a *confirmed* or *validated* exoplanet. Steps such as high-precision RV measurements (Borucki et al. 2010; Koch et al. 2010b; Dunham et al. 2010; Jenkins et al. 2010a; Latham et al. 2010) or transit timing variations (TTVs; Holman et al. 2010; Lissauer et al. 2011a) are used when practical. When such methods cannot be used to *confirm* an exoplanet, an extensive analysis of spacecraft and ground-based data may allow *validation* of an exoplanet by showing that the planetary interpretation is at least 100 times as probable as a false positive (Torres et al. 2011; Lissauer et al. 2011a). This paper does not attempt to promote the candidates discussed herein to validated or confirmed exoplanets, but rather documents the full set of current candidates and the many levels of steps toward eventual validation, or in some cases, rejection as a planet that have been taken.

There are two general causes of false positive events in the *Kepler* data that must be evaluated and excluded before a candidate planet can be considered a valid discovery: (1) statistical fluctuations or systematic variations in the time series and (2) astrophysical phenomena that produce similar signals. A sufficiently high detection threshold (i.e., 7.1σ) was chosen such that the totality of data from Q0 through Q5 (end date JD 2455371.170) provides an expectation of fewer than one false positive event due to statistical fluctuations over the ensemble of all stars for entire mission duration. Similarly, systematic variations in the data have been interpreted in a conservative manner and should result in false positives only rarely. However, astrophysical phenomena that produce transit-like signals are common.

2.2.1. Search for False Positives in the Output of the Data Pipeline

The Transiting Planet Search (TPS) pipeline searches through each systematic error-corrected flux time series for periodic sequences of negative pulses corresponding to transit signatures. The approach is a wavelet-based, adaptive matched filter that characterizes the power spectral density (PSD) of the background process yielding the observed light curve and uses this time-variable PSD estimate to realize a pre-whitening filter and whiten the light curve (Jenkins 2002; Jenkins et al. 2010c, 2010d). TPS then convolves a transit waveform, whitened by the same pre-whitening filter as the data, with the whitened data to obtain a time series of single event statistics. These represent the likelihood that a transit of that duration is present at each time step. The single event statistics are combined into multiple event statistics by folding them at trial orbital periods ranging from 0.5 days to as long as one quarter (∼93 days) of a spacecraft year. Every quarter year, the spacecraft must be rotated 90° to keep the solar panels pointed at the Sun. This rotation put the images of the stars on a different set of detectors and resets the photometric values. Automated identification of candidates with periods longer than one quarter will be done by the pipeline in the coming months, but is currently done by ad hoc methods. The ad hoc methods produced many of the *Kepler*-Objects-of-Interest (KOIs) with numbers larger than 1000, but might cause a bias against candidates with periods longer than one quarter. For a more comprehensive discussion of the data analysis, see Wu et al. (2010) and Batalha et al. (2010b).

After automatic identification with TPS or ad hoc detection of longer period candidates, the light curves of potential planet candidates were modeled and examined by eye to determine the gross viability of the candidate. If the potential candidate was not an obvious variable star or EB showing significant ellipsoidal variation, the candidate was elevated to KOI status, given a KOI number (see Section 3.1) and was subjected to tests described in the next paragraphs. After passing these tests, the KOI is forwarded to the Follow-up Observation Program (FOP) for various types of observations and additional analysis. See the discussion in Gautier et al. (2010) and S. T. Bryson et al. (2011, in preparation).

Using these estimates and information about the star from the KIC, tests are performed to search for a difference in even- and odd-numbered event depths. If a significant difference exists, this suggests that a comparable-brightness EB has been found for which the true period is twice the period initially determined due to the presence of primary and secondary eclipses. Similarly, a search is conducted for evidence of a secondary eclipse or a possible planetary occultation roughly half-way between the potential transits. If a secondary eclipse is seen, then this could indicate that the system is an EB with the period assumed. However, the possibility of a self-luminous planet (as with HAT-P-7; Borucki et al. 2009) must be considered before dismissing a candidate as a false positive.

Many false positives due to background eclipsing binaries (BGEBs) are not detected by the pipeline techniques described above, for example, if their secondary transit signals are so weak that they are lost in the noise. The term "eclipsing binaries," as distinct from BGEBs, are gravitationally bound, multi-star targets, and are usually detected by the secondary eclipse or RV observations. To detect BGEBs, a very sensitive validation technique is used on all candidates to determine the relative position of the image centroid during and outside of the transit epoch. The shift in the centroid position of the target star measured in and out of the transits must be consistent with that predicted from the fluxes and locations of the target and nearby stars. (See S. T. Bryson et al. 2011, in preparation.) In particular, a post-processing examination uses an average difference image formed by subtracting the pixels during transit from the pixels out of transit. A pixel response function fit to this difference image provides a direct sub-pixel measurement of the transit source location on the sky (Torres et al. 2011). When the measured position of the transit source does not coincide with the target star, the most common cause will be a BGEB false positive, although for strongly blended targets in the direct image, further analysis is necessary to support this rejection. This analysis of centroid motion is capable of identifying BGEBs as close as about 1 arcsec to the target star in favorable circumstances, even with *Kepler*'s 4 arcsec pixel scale.

Centroid analysis is conducted for each candidate that is unsaturated in the *Kepler* observations, and follow-up observations by adaptive optics (AO) and speckle imaging of the area near the target star are carried out for many candidates. AO observations in the infrared were conducted at the 5 m at Palomar Observatory and the 6.5 m at the MMT with ARIES; speckle observations were obtained at the WIYN 3.5 m telescope. However, the area behind and immediately surrounding the star can conceal a BGEB that could imitate a candidate signature. The area that could conceal an EB varies with brightness of the target star because of photon noise limitations to AO and speckle searches, but is of order 1 arcsec^2. Model estimates of the a priori probability that an EB is present in the magnitude range that could mimic the transit signal range from 10^{-6} to 10^{-4}. Thus, the estimated number of target star locations that might

have an EB too close to the star to be detected by AO or speckle imaging is 0.15–15 based on observations of 150,000 stars.

A much more comprehensive and intensive analysis has been done for the candidates listed here than was done for the data released in 2010 June (Borucki et al. 2011). Consequently, the fraction of the candidates that are false positives in the active candidate list should be substantially smaller than the earlier estimate.

2.2.2. Estimate of False Positive Rate

While many of the candidates have been vetted through the steps described above, the process of determining the residual false positive fraction for *Kepler* candidates at various stages in the validation process has not proceeded far enough to make good quantitative statements about the expected true planet fraction, or reliability, of the released list. However, we can make rough estimates of the quality of the vetting that the KOIs have had. Several groups of KOIs in Table 2 are distinguished by the FOP ranking flag. These groups have had different levels of scrutiny for false positives and will therefore have different expectations for reliability.

KOIs with ranking of 1 are validated and published planets with expected reliability above 98%. We are reluctant to state a higher reliability since unforeseen issues have led to retractions of apparently well-established planets in other planet detection programs.

KOIs with rankings of 2 and 3 have been subject to thorough analysis of their light curves to look for signs of EB origin, analysis of centroid motion to detect BGEBs confused with their target stars, and varying degrees of spectroscopic and imaging follow-up observation from ground- and space-based observatories. These analyses and follow-up observations are generally sufficient to eliminate many stellar mass objects at or near the location of the target star as the source of the transit signal. A ranking of 2 means that none of the results argued against the planet interpretation. A ranking of 3 means that some of the results were suspicious enough to warrant caution but did not unambiguously rule out the planet interpretation. The criteria are subjective and are not meant to be quantitative. The main sources of unreliability, false positives among the rank 2 and 3 KOIs are likely to be from BGEBs with angular separation from the target star too small to be detected by our centroid motion analysis, grazing eclipses in binary systems, and eclipsing stars in hierarchical multiple systems where transits by stellar companions and giant planets dilute the light of other system components. Note that spectroscopy, even at low signal to noise such as the reconnaissance spectra we are pursuing, easily rules out grazing eclipsing binaries, as they would show RV variations of tens of km s^{-1}. However, those KOIs in Table 2 without a flag = 1 in the FOP column did not have such spectroscopy, leaving open the possibility of such grazing eclipsing binaries.

For bright unsaturated stars with $K_p \leqslant 11.5$ and transit depths strong enough to provide overall detection significances of 20σ and more, the minimum angular separation for the current centroid motion analysis is about 1 arcsec. This limit becomes significantly larger for fainter stars and/or low-amplitude transit signals associated with smaller planets. For these signal levels, the transit significance of $\sim 10\sigma$ supports a centroid motion analysis constraint on the inner detection limit of about 3 arcsec. These minimum detection angles of 1–3 arcsec are quoted as 3σ angles beyond which high confidence of discriminating against BGEBs exists. High-resolution imaging provided additional reduction of the effective minimum detection angle for about 100 of the rank 2 KOIs. We expect 10% of the BGEBs to remain in the rank 2 list. KOIs were given a rank of 3 when the centroid motion analysis or follow-up spectroscopy was ambiguous so that the KOI could not be definitely declared a false positive. We estimate that as many as 30% false positives could remain among the rank 3 KOIs.

About 12% of star systems in the solar neighborhood are found to be triple, or of higher multiplicity, hierarchical systems (Raghavan et al. 2010), so a similar fraction is expected to appear in the *Kepler* target list. Only a small percentage of the hierarchical systems will produce eclipses that are seen by *Kepler* and many of these signals can be identified as binary star eclipses by examination of their light curves. From the rare occurrence rate of EBs and also the rare occurrence rate of triple-star systems, the fraction of KOIs that are triple-star systems with an EB is expected to be less than 5%.

A potentially more frequent type of misidentification in a hierarchical system is a planet transiting in a binary system. If the double nature of the star system is not identified, dilution of the planetary transit by the second star will result in miscalculation of the planet size. Raghavan et al. (2010) give the binary star system fraction as 34%, but little is yet known about the frequency of planets in binary systems and, again, only a small fraction of planets in binary systems will transit because the orbital planes of the planets are expected to be coplanar with the orbital plane of the stars. Adopting the Raghavan et al. occurrence rate of binary stars and assuming that the typical number of planets per star system does not depend on the multiplicity of the system, we expect that up to 34% of the KOIs represent planets of larger radius than indicated in Table 2. The distribution of the amounts of dilution cannot be easily determined as it depends on two effects, namely, the distribution of the ratio of star brightnesses and the distribution of planet sizes that transit one (or the other) of the two stars in the binary system. Estimating these planet-transit effects in binary systems requires knowledge of the systematic dependence of planet size on orbital distance, a chicken-and-egg problem that we cannot easily resolve at present. For binaries in which the transiting planet orbits the primary star, the dilution will be less than 50% flux. But for binaries in which the transiting object (planet or star) orbits the fainter secondary star, the transiting object's radius can be arbitrarily larger than that stated in Table 2.

Considering all sources of remaining false positives, we expect the list of rank 2 KOIs to be >80% reliable and the rank 3 list to be >60% reliable. A careful assessment of false positive scenarios, especially background and gravitationally bound eclipsing binaries and planets, suggests that 90%–95% of the *Kepler* planet candidates are indeed true planets (Morton & Johnson 2011). This agrees with our best estimates.

Rank 4 KOIs have had scant examination of their light curves and no follow-up observation and were therefore subject only to centroid motion analysis. We expect the reliability of rank 4 KOIs to be similar to that of rank 3.

2.2.3. Development of a Model to Estimate the Probability of an EB Near the Position of a Candidate

Low-mass planets, especially those in long-period orbits within the HZ, have low-amplitude RV signal levels that are often too small to be confirmed by current Doppler observation capabilities. Consequently, validation must be accomplished by the series of steps outlined above. An estimate is also made of the probability that an EB is present that is too near the target star

Table 1
Host Star Characteristics

KOI	KIC	K_p (mag)	CDPP (ppm)	R.A. (hr)	Decl. (deg)	T_{eff} (K)	log(g) (cgs)	R^* (R_{sun})	M (M_{sun})	T_{eff} Flag
1	11446443	11.338	14	19.12056	49.3164	5713	4.14	1.50	1.14	
2	10666592	10.463	21.9	19.48315	47.9695	6577	4.32	1.34	1.36	1
3	10748390	9.147	97.8	19.84729	48.0809	4628	4.53	0.76	0.71	1
4	3861595	11.432	126	19.62377	38.9474	6054	4.41	1.08	1.11	
5	8554498	11.665	20.2	19.31598	44.6474	5766	4.04	1.73	1.18	
7	11853905	12.211	71.2	19.04102	50.1358	5701	4.35	1.16	1.08	
10	6922244	13.563	58.6	18.75254	42.4511	6164	4.44	1.05	1.12	
12	5812701	11.353	82	19.83025	41.0110	6419	4.26	1.32	1.17	
13	9941662	9.958	10.4	19.13141	46.8684	8848	3.93	2.44	1.83	
17	10874614	13.000	38.6	19.78915	48.2399	5724	4.47	0.91	0.91	1

Notes. All parameters are from the *Kepler* Input Catalog (KIC) except where T_{eff} Flag = 1 indicates that no parameters were available in the KIC. In which case T_{eff}, log(g), and R^* are derived as noted.
Key:
KOI *Kepler*-Object-of-Interest number
KIC *Kepler* Input Catalogue Identifier
K_p *Kepler* magnitude
CDPP 6 hr Combined Differential Photometric Precision from Quarter 3
R.A. Right ascension (J2000)
Decl. Declination (J2000)
T_{eff} Effective Temperature of host star as reported in the KIC. If T_{eff} Flag = 1, then T_{eff}, log(g), R are derived using KIC $J - K$ color and linear interpolation of luminosity class V stellar properties of Schmidt-Kaler (1982).
log(g) Surface gravity reported by KIC. If T_{eff} Flag = 1, then log(g) is based on $J - K$ interpolation.
R^* Stellar radius reported by KIC. If T_{eff} Flag = 1, then R is based on $J - K$ interpolation.
M Stellar mass derived from log(g) and stellar radius.

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)

to detect by AO, speckle imaging, or centroid motion. The area number density (number per solid angle) of EBs is calculated based on the assumption that the number of EBs to the number of background stars is constant near the position of each target star. Because the area number density varies rapidly with Galactic latitude and because the *Kepler* field of view (FOV) covers over 10° of latitude, predictions of the EB density also vary greatly over the FOV. Consequently, a model was constructed to estimate the probability per square arcsec that an EB is present in the magnitude range that would provide a signal with an amplitude similar to that of the candidate and at the position of each target star. The model is based on the fraction of stars observed by *Kepler* to be binary (Prsa et al. 2011), and it uses the number and magnitude distributions of stars from the Besancon model (Drimmel et al. 2003) after correction from the V band to the *Kepler* passband. The value of the probability that there is a BGEB at the location of the target star is listed in Table 2 for each candidate.

3. RESULTS

The characteristics of the host stars and the candidates are summarized in Tables 1 and 2, respectively. A total of 1235 KOIs were found in the Q0 through Q2 data. Table 3 provides short notes on many of these KOIs. Table 4 lists the 511 candidates considered to be false positives; comments are included. The false positives have been removed from the list of candidates in Table 2 and are not used in the distributions discussed here. The 15 candidates with a diameter over twice that of Jupiter, and thus larger than late M dwarf stars, were also removed from discussion. This leaves a total of 1235 candidates: 18 single-transit candidates, 15 candidates greater than twice the size of Jupiter, and 1202 candidates for consideration in this discussion.

To provide the most accurate predictions for future observations, the values for the epoch and orbital period given in Table 2 are derived from all data currently available to the *Kepler* team, i.e., data obtained through Q5 (from JD 2455276.481 through JD 2455371.170) were used. For some candidates, reconnaissance spectra were taken with moderate exposures to look for double- and single-lined binaries. They are most useful in finding outliers for the stellar temperatures and log g listed in the KIC. AO and speckle observations were taken to check for the presence of faint nearby stars that could be BGEBs or that could dilute the signal level. Flags also indicate the particularly interesting candidates for which RV measurements of extremely high precision (~2 m s^{-1}) or high-precision (~10 m s^{-1}) observations were obtained. The last column of Table 2 indicates whether a note is available about that candidate in Table 3. For consistency, all values of the stellar parameters are derived from the KIC.

3.1. Naming Convention

To avoid confusion in naming the target stars, host stars, planetary candidates, and confirmed/validated planets, the following naming convention has been used. *Kepler* stars are referred to as KIC NNNNNNN (with a space between the "KIC" and the number), where the integer refers to the ID in the *Kepler* Input Catalog archived at MAST. Confirmed planets are named *Kepler* followed by a hyphen, a number for the planetary system, and a letter designating the first, second, etc., confirmed planet as "b," "c," etc., for example, Kepler-4b. Candidates are labeled "KOI" followed by a decimal number. The two digits beyond the decimal provide identification of the candidates when more than one is found for a given star, e.g., KOI NNN.01, KOI NNN.02, KOI NNN.03, etc. For example, KOI 377.03, the third transit candidate identified around star KOI 377, became

Table 2
List of Planetary Candidates and Their Characteristics

KOI	Dur (hr)	Depth (ppm)	S/N	t_0 (BJD−2454900)	t_0_unc	Period (days)	P_unc	a/R^*	a/R^*_unc	r/R^*	r/R^*_unc	b	b_unc	R_p (R_{Earth})	SMA (AU)	T_{eq} (K)	EB Prob	V	FOP	N
1.01	1.7952	14174	2062	55.76258	0.00004	2.4706131	0.0000004	8.519	0.082	0.12429	0.00029	0.816	0.067	20.3	0.037	1603	1.4E-06	1		1
2.01	3.9107	6716	2413	54.35781	0.00005	2.2047355	0.0000004	4.152	0.041	0.07931	0.00012	0.51	0.1	11.6	0.037	1743	2.4E-06	1		1
3.01	2.3607	4197	328	57.81227	0.00033	4.8878177	0.0000089	16.1	9.1	0.0577	0.0073	0.29	0.86	4.8	0.05	796	2.0E-06	1		1
4.01	2.3866	1193	136	90.5261	0.00055	3.84937	0.000014	10	24	0.034	0.015	0.7	1.4	4.0	0.05	1242	...	3		1
5.01	2.0326	951	263	65.9735	0.00025	4.7803247	0.0000058	7.3	2.2	0.03707	0.0002	0.91	0.27	7.0	0.059	1376	...	3		1
7.01	3.6234	741	231	56.61126	0.00041	3.213682	0.000011	3.94	0.56	0.02911	0.00069	0.86	0.23	3.7	0.044	1290	9.5E-06	1		1
10.01	3.2860	9390	237	54.11809	0.00062	3.522297	0.00008	8.15	0.34	0.09138	0.00071	0.53	0.21	10.5	0.047	1287	6.5E-06	1		1
12.01	7.4343	9253	604	79.59772	0.00038	17.855038	0.000038	19.9	0.025	0.0874	0.0001	0.0003	...	12.6	0.141	868	9.4E-06	3	NoObs	1
13.01	3.2029	4644	1147	53.56498	0.00012	1.7635892	0.0000014	4.51	0.2	0.07695	0.00043	0.26	0.24	20.5	0.035	3257	1.8E-06	2	2,3	1
17.01	3.9011	10738	724	54.48575	0.00007	3.2347003	0.0000012	6.9639	0.0036	0.09467	0.00004	...	...	9.4	0.041	1192	1.6E-05	1		1

Notes.
Key:

KOI — *Kepler*-Object-of-Interest number. † indicates that this KOI was detected on the basis of a single transit with the period derived from the transit duration and stellar radius.

Dur — Transit duration, first contact to last contact.

Depth — Transit depth at center of transit.

S/N — Total S/N of all transits detected. S/N = Depth/(Std∗sqrt(N)), where Std is the standard deviation of all data outside of transits (Q0 through Q5) and N is the total number of measurements inside of all transits.

t_0, t_0_unc — Time of a transit center based on a linear fit to all observed transits and its uncertainty.

Period, P_unc — Average interval between transits based on a linear fit to all observed transits and uncertainty.

a/R^*, a/R^*_unc — Ratio of semimajor axis to stellar radius assuming zero eccentricity, a parameter derived from the light curve and uncertainty.

r/R^*, r/R^*_unc — Ratio of planet radius to stellar radius and uncertainty.

b, b_unc — Impact parameter of the transit and uncertainty. Note, there is a strong co-variance between b and a/R^*.

R_p — Radius of planet in units of $R_{Earth} = 6378$ km.

a — Semimajor axis of orbit based on Newton's generalization of *Kepler*'s third law and the stellar mass in Appendix 1.

T_{eq} — Equilibrium temperature of the planet (see the main text and Appendix 5 for discussion).

EB prob — Probability of BGEB confused with planet's host star (see the text for discussion).

V — Vetting flag
1 Confirmed and published planet.
2 Strong probability candidate, cleanly passes tests that were applied.
3 Moderate probability candidate, not all tests cleanly passed but no definite test failures.
4 Insufficient follow-up to perform full suite of vetting tests.

FOP — Follow-up observation description (to be revised)
1 Reconnaissance spectra taken.
2 Adaptive optics observations taken.
3 Speckle observations taken.
4 10 m s^{-1} RV spectra taken.
5 2 m s^{-1} RV spectra taken.
NoObs No observations yet taken.

N — Notes flag. A "1" indicates a note on this KOI or its host star in Appendix 3.

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)

Table 3
Notes to Table of Planet Candidate Characteristics

KOI	Note
1.01	TrES-2; O'Donovan et al. (2006)
2.01	HAT-P-7; Pál et al. (2008)
3.01	HAT-P-11b; Bakos et al. (2010)
4.01	Rapid rotator $V_{rot} = 40$ km s^{-1}
5.01	Double star; 0.″16 NE; delta_m = 3.1 at 692 nm
7.01	Kepler-4b; Borucki et al. (2010)
10.01	Kepler-8b; Jenkins et al. (2010a)
12.01	Marginally saturated
13.01	Double star; 0.″8 E; delta_m = 0.4 mag at 692 nm
17.01	Kepler-6b; Dunham et al. (2010)
18.01	Kepler-5b; Koch et al. (2010b)
44.01	Variable transit depths
51.01	Light curve has spot/rotation modulation
63.01	Radial velocity variations have a dispersion 23 m s^{-1}
64.01	May be an F-M binary
69.01	Saturated. Double star; 0.″05 NW; delta_mag = 1.4 mag
72.01	Kepler-10b; Batalha et al. (2011)
97.01	Kepler-7b; Latham et al. (2010)
99.01	Double star; 4″ SE
100.01	Rapid rotator; $V_{rot} = 35$ km s^{-1}
102.01	Double star; 2.″5 SW
112.01	Double star; 0.″09; delta_m = 2.7 at 692 nm
117.02	Possible APO
117.03	Possible APO
119.01	Possible SB1
131.01	Possible APO
135.01	Centroid analysis clean
144.01	KIC radius likely overestimated
151.01	V-shaped; may be triple system
155.01	Double star; 2″ W
157.01	Kepler-11b; Lissauer et al. (2011a)
157.02	Kepler-11c; Lissauer et al. (2011a)
157.03	Kepler-11d; Lissauer et al. (2011a)
157.04	Kepler-11e; Lissauer et al. (2011a)
157.05	Kepler-11f; Lissauer et al. (2011a)
157.06	Kepler-11g; Lissauer et al. (2011a)
179.01	Double Star; 4″ E
180.01	Variable star
184.01	Odd–even
191.01	Double star; 1″ E
191.02	Possible APO; Double star 1″ E
208.01	Variable star with possible spots
225.01	Possible ellipsoidal variations
226.01	Possible APO
254.01	5% primary transit
256.01	KIC stellar radius may be too large
258.01	V-shaped; Multiple stars 1″ and 2″ E
263.01	Double star; 4″ E
268.01	Multiple Stars: 2″ S and 3″ SE
271.02	Possible Odd–even
274.01	Possible APO
284.01	Double star; 0.″9 E
340.01	Radius large; but log g may be too low in the KIC
377.01	Kepler-9b; Holman et al. (2010)
377.02	Kepler-9c; Holman et al. (2010)
377.03	Kepler-9d; Torres et al. (2011)
531.01	Strange light curve; worth follow-up
607.01	Odd light curve; worth follow-up
687.01	Varying depths; possible encroaching companion
741.01	Slight V shape and deep; no APO
774.01	Possible occultation
961.01	Short duration, under sampled transit
962.01	Weak transit signal; possible low radius planet
968.01	Not convincing transit
972.01	Pulsating star
973.01	Possible APO; poor light curve

Table 3
(Continued)

KOI	Note
976.01	V-shaped; poor fit
977.01	Phase-correlated variations; saturated
978.01	Possibly spurious
981.01	V-shaped; saturated
984.01	V-shaped
992.01	Poor fit to light curve
993.01	Possible APO
994.01	Possible APO
998.01	Eccentric EB
1063.01	V-shaped; large planet radius (2.1 RJ)

Notes.
Key:

APO	Active pixel offset. The pixel that actually dims during a transit is offset from the position of the target star implying a background variable star.
Double star	There is within 4″ an object evident in images that has not been ruled out as the source of the transit.
V-shaped	The transit light curve is "V" shaped, a possible indication of an EB.
Odd–even	Transit depths are alternately deeper and shallower, an indication of an EB.
Occultation	Evidence of secondary eclipse, implying possible EB or self-luminous planet.
SB1	Spectroscopic binary. RV varies by over 1 km s^{-1} in low S/N reconnaissance spectra. Double lines not seen.
SB2	Spectroscopic binary. Double lines seen in spectrum.

Kepler-9d after validation as a planet (Torres et al. 2011). KOI numbers are always cross-referenced to a KIC ID. For a multi-candidate system these digits beyond the decimal indicate the order in which the candidates were identified by the analysis pipelines and are not necessarily in order of orbital period. It should be noted that the KOI list is not contiguous and not all integers have an associated KOI.

3.2. Statistical Properties of Planet Candidates

We conducted a statistical analysis of the 1202 candidates to investigate the general trends and initial indications of the characteristics of the planetary candidates. The list of candidates was augmented with known planets in the field of view. In particular, TrES-2, HAT-P7b, HAT-P11b, (Kepler-1b, -2b, -3b, respectively), Kepler-4b–8b (Borucki et al. 2010; Koch et al. 2010b; Dunham et al. 2010; Latham et al. 2010; Jenkins et al. 2010a). Kepler-9bcd (Holman et al. 2010; Torres et al. 2011), Kepler-10b (Batalha et al. 2011), and Kepler-11b–g (Lissauer et al. 2011a) were included. However, one candidate identified by a guest observer (KOI 824.01) is included in the list of candidates but is not used in the graphs and statistics because it was not in the range of parameters chosen for the search. As noted above, not all candidates appearing in Table 2 were used in the statistical analysis or in the graphical associations shown in the figures: specifically, candidates greater than twice the size of Jupiter, those that showed only one transit in the Q0/Q2 data but no others in the succeeding observations, and those orbiting stars larger than 10 solar radii or with temperatures in excess of 9500 K were excluded. Comparisons are limited to orbital periods of $\leqslant$138 days. The figures are indicative of the properties and associations of candidates with various parameters, but are not meant to be definitive.

The readers are cautioned that the sample is affected by many poorly quantified biases. Obviously some of the released

Table 4
Very Probable False Positives

KOI	*Kepler* ID	t_0 (BJD−2454900)	Period (days)	Depth (ppm)	S/N	Comment
6.01	3248033	66.69954	1.334103	397	97	APO Binary
8.01	5903312	54.70223	1.160154	399	41	APO Binary
9.01	11553706	68.06724	3.719813	3423	380	APO Binary
11.01	11913073	104.65803	3.748075	547	65	APO Binary
14.01	7684873	104.53055	2.947317	302	59	Rapid rotator; $V_{\rm rot} = 90$ km s^{-1}; Secondary eclipse
15.01	3964562	68.25804	3.012481	1599	301	APO Binary
16.01	9110357	66.40566	0.895298	1527	283	APO Binary
19.01	7255336	66.93003	1.203197	2472	92	Binary, Odd–even
21.01	10125352	54.97329	4.288459	3127	246	Binary
23.01	9071386	69.86191	4.693309	14756	1443	SB1; 18 km s^{-1} radial velocity amplitude; secondary eclipse in light curve

Notes.
Key:

t_0	Time of the transit center based on a linear fit to all observed transits and its uncertainty.
Period	Average interval between transits based on a linear fit to all observed transits and uncertainty.
APO	Active pixel offset. The pixel that actually dims during a transit is offset from the position of the target star implying a background variable star.
Double star	There is within 4″ an object evident in images that has not been ruled out as the source of the transit.
V-shaped	The transit light curve is "V" shaped, a possible indication of an EB.
Odd–even	Transit depths are alternately deeper and shallower, an indication of an EB.
Occultation	Evidence of secondary eclipse, implying possible EB or self luminous planet
SB1	Single-line EB star. RV varies by over 1 km s^{-1} in low S/N reconnaissance spectra. Double lines not seen.
SB2	Double-line EB. Double lines seen in spectrum.

(This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content.)

candidates could be false positives, but other characteristics such as stellar radius, magnitude, noise spectrum, and analysis protocols can all play significant roles in the statistical results. Nevertheless, the large number of candidates provides interesting, albeit tentative, associations with stellar properties. No correction is made to the frequency plots due to the linearly decreasing probability of a second transit occurring during the Q0 through Q2 period. This correction is not needed because data for following quarters were used to calculate the epochs and periods for all candidates that showed at least one transit in the Q0 through Q2 period and at least one in the subsequent observations. In the figures below, the distributions of various parameters are plotted and compared with values in the literature and those selected from the Extrasolar Planets Encyclopedia[31] (EPE; values as of 2010 December 7). We consulted the literature to identify those planets discovered by the RV method and excluded those discovered by the transit method. This step avoids biasing the RV-discovered planets with the short-period planets that are often found by the transit method.

The results discussed here are primarily based on the observations of stars with $K_{\rm p} < 16$, with effective temperature below 9500 K and with size less than 10 times the solar radius. The latter condition is imposed because the photometric precision is insufficient to find Jupiter-size and smaller planets orbiting stars with 100 times the area of the Sun. Stellar parameters are based on KIC data. The function of the KIC was to provide a target sample with a high fraction of dwarf stars that are suitable for transit work and to provide a first estimate of stellar parameters that is intended to be refined spectroscopically for KOI targets at a later time. Although post-identification reconnaissance spectroscopic observations have been made for more than half of the stars with candidates, it is important to recognize that some of the characteristics listed for the stars are still uncertain, especially surface gravity (i.e., log g) and metallicity ([M/H]). The errors in the stellar diameters can reach 25%, with proportional changes to the estimated diameter of the candidates.

In Figure 1, the stellar distributions of magnitude and effective temperature are given for reference. In later figures, the association of the candidates with these properties is examined.

It is clear from the left panel in Figure 1 that most of the stars monitored by *Kepler* have temperatures between 4000 and 6500 K; they are mostly late F, G, and K spectral types. Because of their faintness, only 2510 stars cooler than 4000 K (i.e., dwarf stars of spectral type M) were monitored. Although cooler stars are more abundant, hotter stars are the most frequently seen for a magnitude-limited survey of dwarfs.

The selection of target stars was purposefully skewed to enhance the detectability of Earth-size planets by choosing those stars with an effective temperature and magnitude that maximized the transit S/N (Batalha et al. 2010b). The step decrease seen in the right-hand panel of Figure 1 at *Kepler* magnitude ($K_{\rm p}$) equals 14.0 and the turnover near $K_{\rm p} = 15.5$, seen in the right-hand panel of Figure 1, are due to the selection of only those stars in the FOV that are bright enough and small enough to show terrestrial-size planets. After all available bright dwarf stars were chosen for the target list, many target slots remained, but only stars fainter than $K_{\rm p} = 14$ were available (Batalha et al. 2010b). From the fainter stars the smallest stars are given preference. At the lower left of the right-hand chart, the bin size has been increased to show the small number of candidates brighter than $K_{\rm p} = 9$. In the following figures, the bias introduced by the selection of stellar size and magnitude distributions must always be considered.

As noted in Borucki et al. (2011), the results shown in Figure 2 imply that small candidate planets are much more common than large candidate planets. Of the 1202 candidates considered for the analysis, 74% are smaller than Neptune ($R_{\rm p} = 3.8\,R_{\oplus}$). Table 6 shows the observed distribution and the definition of sizes used throughout the paper for these 1202 candidates.

[31] Extrasolar Planet Encyclopedia; http://exoplanet.eu/.

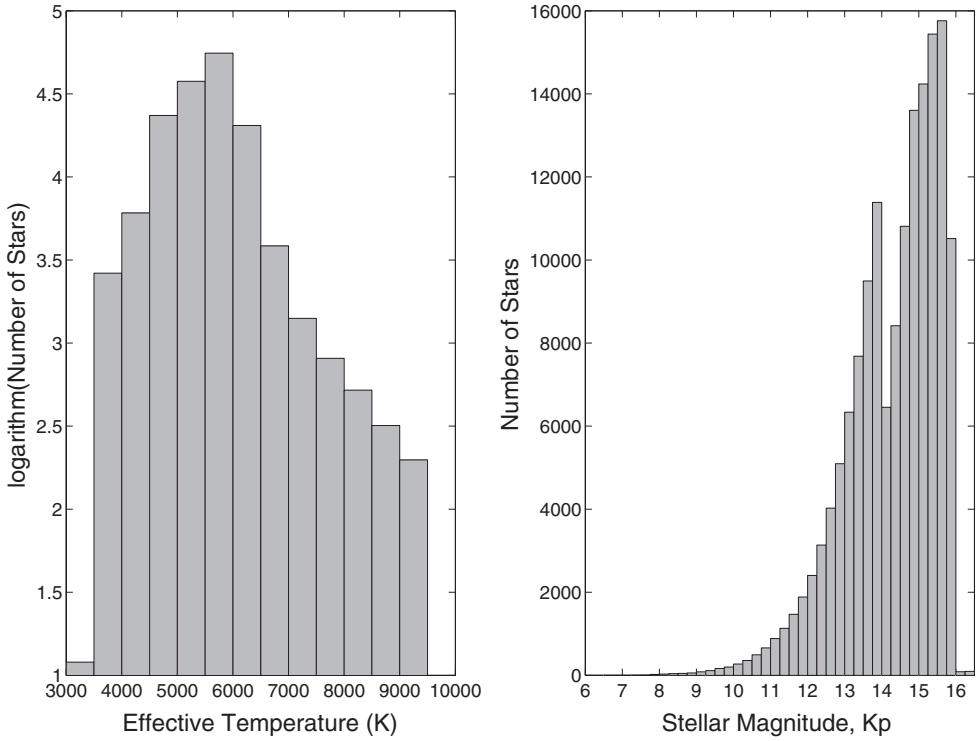

Figure 1. Distributions of effective temperature and magnitude for the stars observed during Q2 and considered in this study. Bin size for the left panel is 500 K. The bin size for the right-hand panel is 1 mag from 6 to 9 and 0.25 mag from 9 to 16.5.

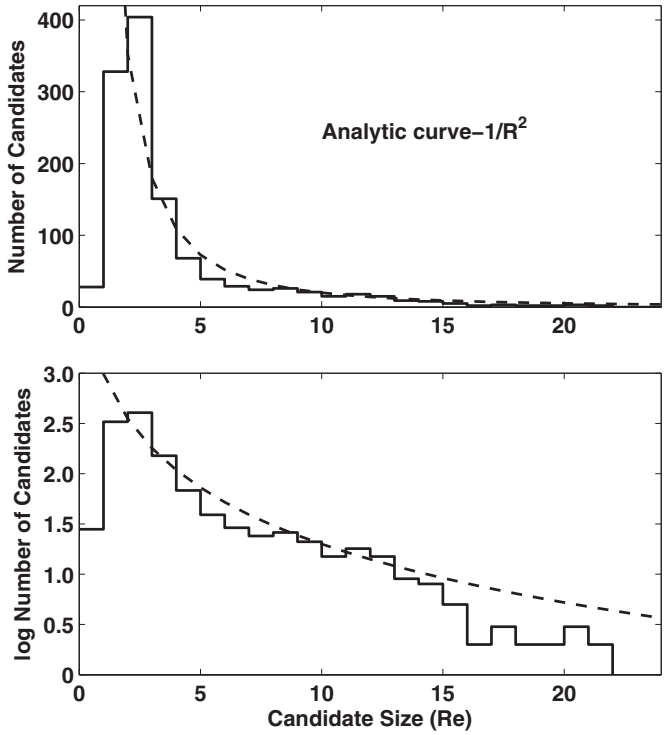

Figure 2. Size distribution of the number of *Kepler* candidates vs. planet radius (R_p) (upper panel). The logarithm of the number of candidates is presented in the lower panel to better show the tail of the distribution. Bin sizes in both panels are 1 $R_\oplus$.

The dashed curve in both panels of Figure 2 represents a $(1/{R_p}^2)$ dependence of the number of candidates on candidate radius, i.e., dN/dr scales as the reciprocal of the cube root of R_p for $2\,R_\oplus < R_p < 15\,R_\oplus$. The data shown here are restricted to orbital periods $\leqslant$138 days. Because it is much easier to detect larger candidates than smaller ones, this is a robust result that implies the frequency of candidates decreases with the area of the candidate, assuming that the false positive rate, completeness, and other biases are independent of candidate size for candidates larger than two Earth radii. However, the current survey is not complete, especially for the fainter stars, smallest candidates, and long orbital periods, and further observations could influence the distribution.

Figure 3 presents scatter plots showing the observed relative size of individual candidates versus orbital period, semimajor axis, stellar temperature, and candidate temperature. The values on the abcissa are limited to show only the most populous range. Outliers can be found in Table 2. The upper left panel shows a concentration (in log–log space) of candidates for orbital periods between 3 and 30 days and sizes between 1 and 4 $R_\oplus$. The upper right panel shows a similar concentration. Both of them show a nearly empty area to the lower right that likely represents the lack of small candidates caused by the lower detectability of small candidates in long-period orbits.

All panels in Figure 3 show a scarcity of candidates with radius R_p smaller than 1 $R_\oplus$. The paucity of small candidates at even the shortest orbital periods could be due to incompleteness for the smaller signals, coupled with the analysis of only a portion of the expected *Kepler* data, and higher than expected noise levels. These effects could mask a real dependence of number on size. The modestly higher noise levels than those anticipated are thought to follow primarily from an underestimate of intrinsic stellar noise and are the topic of an on-going study.

Figure 4 expands that portion of the lower right panel to emphasize those candidates with estimated radiative equilibrium temperatures in the range of liquid water at a pressure of 1 bar.

The HZ is often defined to be that region around a star where a rocky planet with an Earth-like atmosphere could have a surface temperature between the freezing point and boiling

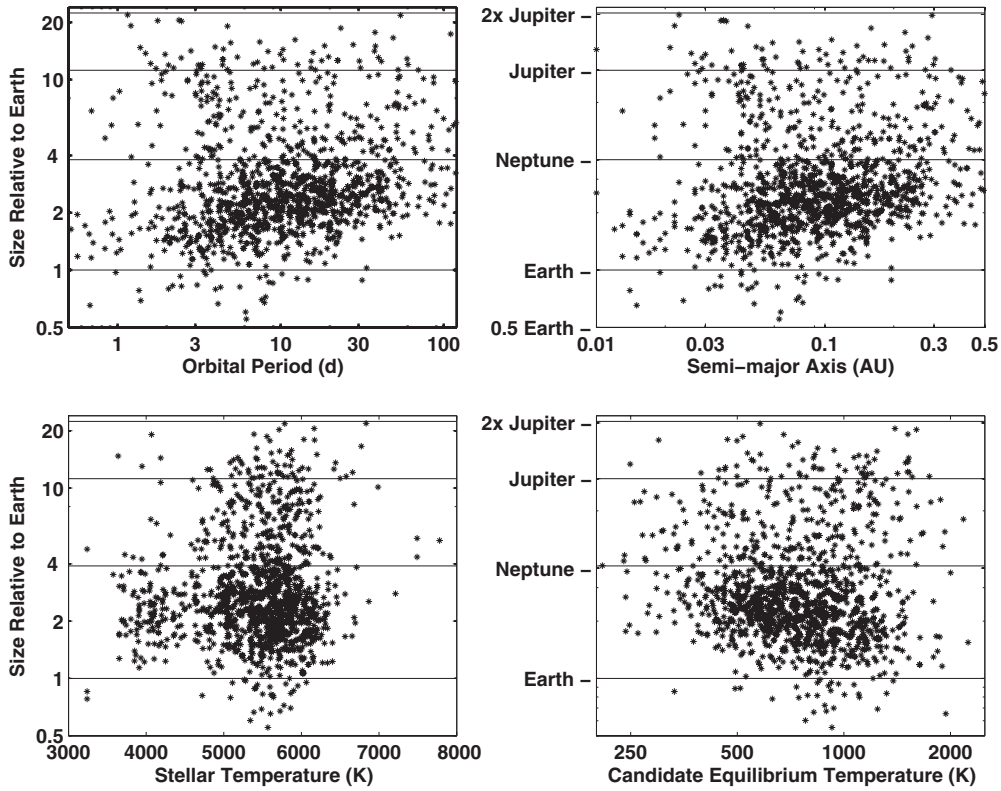

Figure 3. Candidate size vs. orbital period, semimajor axis, stellar temperature, and candidate equilibrium temperature (T_{eq} was derived by assuming an even distribution of heat from the day to night side of the planet (e.g., a planet with an atmosphere or a planet with rotation period shorter than the orbital period), and the planet and star act as blackbodies in equilibrium: $T_{eq} = T_*(R_*/2a)^{1/2}[f(1 - A_B)]^{1/4}$, where T_* and R_* are the effective temperature and radius of the host star, the planet at distance a with a Bond albedo of A_B, and f is a proxy for atmospheric thermal circulation. The Bond albedo, A_B, is the fraction of total power incident on a body scattered back into space which we assume to be 30% and $f = 1$ indicates full thermal circulation.). Uncertainties in candidate size are mostly due to the uncertainty in stellar sizes, i.e., approximately 25%. Horizontal lines mark ratios of candidate sizes for Earth-size, Neptune-size, and Jupiter-size relative to Earth-size.

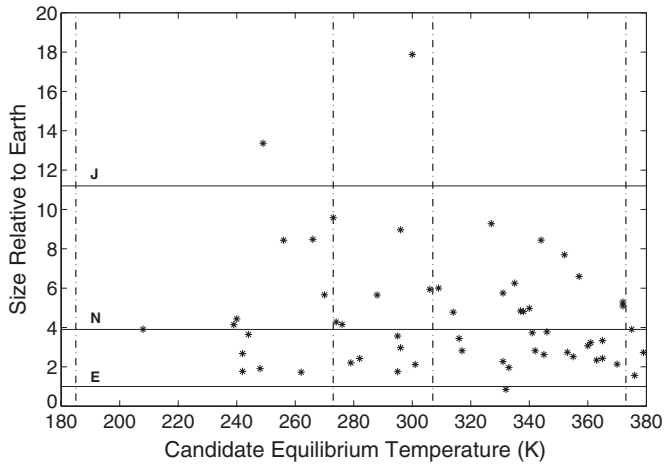

Figure 4. Candidate sizes and estimated radiative equilibrium temperatures (T_{eq}) centered on the HZ temperature range. The dotted lines at 273 K and 373 K bracket the range of surface temperatures allowing water to exist as a liquid at one atmosphere of pressure. Uncertainties are discussed in the text.

point of water, or analogously the region receiving roughly the same insolation as the Earth from the Sun (Kasting et al. 1993; Rampino & Caldeira 1994; Heath et al. 1999; Joshi 2003; Tarter et al. 2007). The surface temperature range for HZs is likely to include radiative equilibrium temperatures well below 273 K because of warming by any atmosphere that might be present. For example, the greenhouse effect raises the Earth's surface temperature by 33 K and that of Venus by approximately 500 K. Further, the spectral characteristics of the stellar flux vary strongly with T_{eff} and affect both the atmospheric composition and the chemistry of photosynthesis (Heath et al. 1999; Segura et al. 2005; Kaltenegger & Sasselov 2011). Consequently, Figure 4 includes temperatures well below the freezing point of water. The vertical lines at 183 and 307 K delineate the radiative temperature range for which the surface temperature of a rocky planet with an atmosphere similar to that of the Earth is expected to be within the freezing and boiling point of water (J. E. Kasting 2011, private communication).

The calculated equilibrium temperatures shown in Figure 4 are for gray-body spheres without atmospheres. The calculations assume a Bond albedo of 0.3 and a uniform surface temperature. The uncertainty in the computed equilibrium temperatures is approximately 22% because of uncertainties in the stellar size, mass, and temperature as well as the planetary albedo. For planets with an atmosphere, the surface temperature would be higher than the radiative equilibrium temperature.

Within this temperature range, there are 54 candidates present with sizes ranging from Earth-size to larger than that of Jupiter. Table 5 lists the candidates in the HZ. The detection of Earth-size candidates depends on the signal level, which in turn depends on the size of the candidate relative to the size of the star, the number of transits observed, and the combined noise of the star and the instrument. It is important to recognize that the size of the star is generally not well characterized until spectroscopic studies and analysis are completed. In particular, some of the cooler stars could be nearly double the size shown in Table 1 and that some of the candidates could prove to be false positives.

As can be seen in Table 5, there are two candidates with radii less than 1.5 $R_\oplus$ (KOI 314.02 and KOI 326.01) present in the list.

Table 5
Candidates in or near the HZ (sorted by T_{eq})

KOI	K_p (mag)	R_p ($R_\oplus$)	Period (days)	T_{eff} (K)	R^* ($R_\odot$)	T_{eq} (K)	a (AU)
683.01	13.71	4.14	278.12	5624	0.78	239	0.84
1582.01	15.4	4.44	186.38	5384	0.64	240	0.63
1026.01	14.75	1.77	94.1	3802	0.68	242	0.33
1503.01	14.83	2.68	150.24	5356	0.56	242	0.54
1099.01	15.44	3.65	161.53	5665	0.55	244	0.57
854.01	15.85	1.91	56.05	3743	0.49	248	0.22
433.02	14.92	13.37	328.24	5237	1.08	249	0.94
1486.01	15.51	8.43	254.56	5688	0.83	256	0.8
701.03	13.73	1.73	122.39	4869	0.68	262	0.45
351.01	13.8	8.48	331.65	6103	0.94	266	0.97
902.01	15.75	5.66	83.9	4312	0.65	270	0.32
211.01	14.99	9.58	372.11	6072	1.09	273	1.05
1423.01	15.74	4.28	124.42	5288	0.66	274	0.47
1429.01	15.53	4.15	205.93	5595	0.86	276	0.69
1361.01	14.99	2.2	59.88	4050	0.59	279	0.24
87.01	11.66	2.42	289.86	5606	1.14	282	0.88
139.01	13.49	5.65	224.79	5921	0.9	288	0.74
268.01	10.56	1.75	110.37	4808	0.79	295	0.41
1472.01	15.06	3.57	85.35	5455	0.56	295	0.37
536.01	14.5	2.97	162.34	5614	0.84	296	0.59
806.01	15.4	8.97	143.18	5206	0.88	296	0.53
1375.01	13.71	17.88	321.22	6169	1.17	300	0.96
812.03	15.95	2.12	46.19	4097	0.57	301	0.21
865.01	15.09	5.94	119.02	5560	0.73	306	0.47
351.02	13.8	6	210.45	6103	0.94	309	0.71
51.01	13.76	4.78	10.43	3240	0.27	314	0.06
1596.02	15.16	3.44	105.36	4656	0.98	316	0.42
416.02	14.29	2.82	88.25	5083	0.75	317	0.38
622.01	14.93	9.28	155.05	5171	1.17	327	0.57
555.02	14.76	2.27	86.5	5218	0.78	331	0.38
1574.01	14.6	5.75	114.73	5537	0.85	331	0.47
326.01	12.96	0.85	8.97	3240	0.27	332	0.05
70.03	12.5	1.96	77.61	5342	0.7	333	0.35
1261.01	15.12	6.25	133.46	5760	0.9	335	0.52
1527.01	14.88	4.84	192.67	5470	1.31	337	0.67
1328.01	15.67	4.81	80.97	5425	0.72	338	0.36
564.02	14.85	4.97	127.89	5686	0.93	340	0.51
1478.01	12.45	3.73	76.13	5441	0.7	341	0.35
1355.01	15.9	2.81	51.93	5529	0.52	342	0.27
372.01	12.39	8.44	125.61	5638	0.95	344	0.5
711.03	13.97	2.62	124.52	5488	1	345	0.49
448.02	14.9	3.78	43.62	4264	0.71	346	0.21
415.01	14.11	7.7	166.79	5823	1.15	352	0.61
947.01	15.19	2.74	28.6	3829	0.64	353	0.15
174.01	13.78	2.52	56.35	4654	0.8	355	0.27
401.02	14	6.6	160.01	5264	1.4	357	0.59
1564.01	15.29	3.07	53.45	5709	0.56	360	0.28
157.05	13.71	3.23	118.38	5675	1	361	0.48
365.01	11.2	2.34	81.74	5389	0.86	363	0.37
374.01	12.21	3.33	172.67	5829	1.26	365	0.63
952.03	15.80	2.4	22.78	3911	0.56	365	0.12
817.01	15.41	2.1	23.97	3905	0.59	370	0.13
847.01	15.20	5.1	80.87	5469	0.88	372	0.37
1159.01	15.33	5.3	64.62	4886	0.91	372	0.30

The uncertainty in the sizes of these candidates is approximately 25% to 35% due to the uncertainty in size of stars and of the transit depth.

The predicted semi-amplitudes of the RV signals for small candidates such as KOI 314.02 and 326.01 are 1.2 m s^{-1} and 0.5 m s^{-1}, respectively. These RV amplitudes follow from assuming a circular orbit and a density of 5.5 g cm^{-3} for both candidates. RV semi-amplitudes of 1.0 m s^{-1} are at the very limit of what might currently be possible to detect with the largest telescopes and best spectrometers. In principle, RV amplitudes under 1 m s^{-1} could be detected, but there are many impediments to achieving such precision including the surface velocity fields (turbulence) and spots on the rotating surface. In addition, stars with one transiting planet may well harbor multiple additional planets that do not transit, causing additional RV variations. Moreover, these two stars have *V*-band

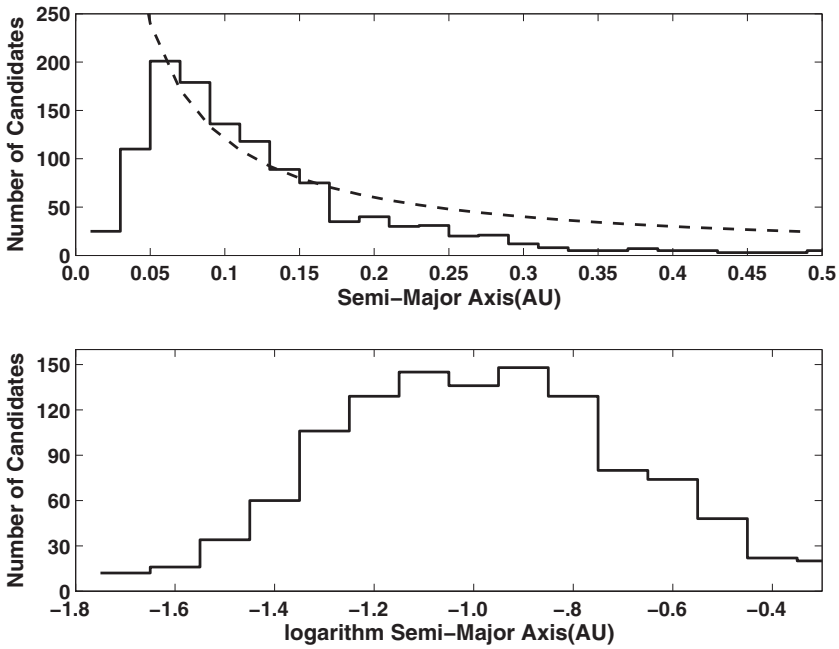

Figure 5. Upper panel: historgrams of the observed number of candidates vs. linear intervals in the semimajor axis. The dashed line shows the relative effect of the orbital alignment probability. Lower panel: the number of candidates vs. logarithmic intervals of the semimajor axis. Bin size is 0.02 AU in the upper panel and 0.1 in the lower panel.

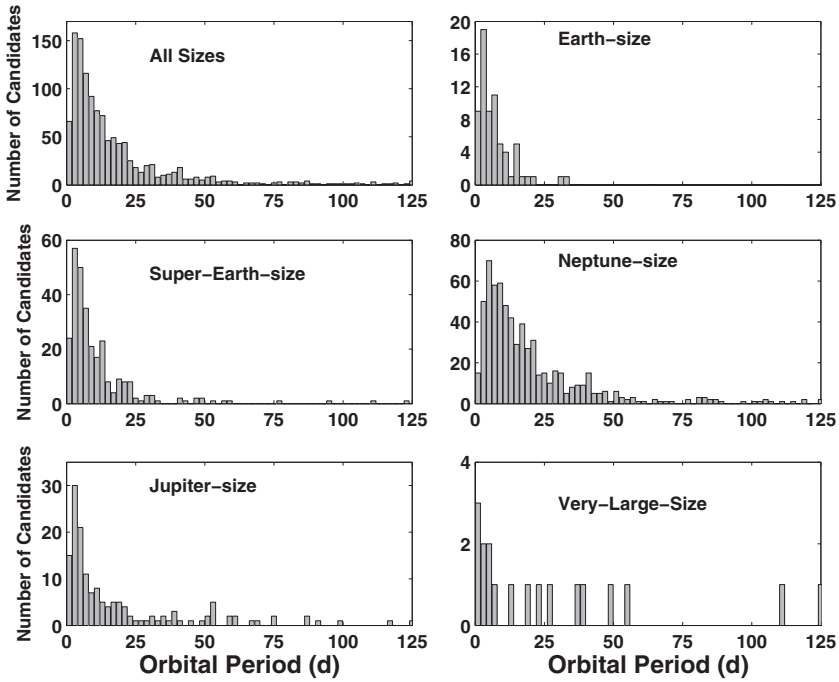

Figure 6. Number of candidates vs. orbital period for several choices of candidate size. Bin size is 2 days. Refer to Table 6 for the definition of each size category.

magnitudes of 14, making it very difficult to acquire sufficient photons in a high-resolution spectrum to achieve the required Doppler precision. Of course, for all of these small planets RV measurements can place firm upper limits to their masses and densities.

In Figure 5, the dependence of the number of candidates on the semimajor axis is examined. For a less than 0.04 AU, it is evident that the distribution is severely truncated. As will be shown in Figure 6, this feature is present in each of the candidate size groups. In the upper panel of Figure 5, an analytic curve shows the expected reduction in the number in each interval due to the decreasing geometrical probability that orbits are aligned with the line of sight. It has been scaled over the range of semimajor axis from 0.04 to 0.5 AU, corresponding to orbital periods from 3 days to 138 days for a solar-mass star. The fact that the fit is fair to poor implies that the intrinsic distribution is not consistent with a simple correction for the orbital alignment probability.

The panels in Figure 6 show that the period distribution of Neptune-size candidates has a less steep slope compared to Jupiter-size candidates in the period range from one week to one month. Because of the large numbers in both samples and

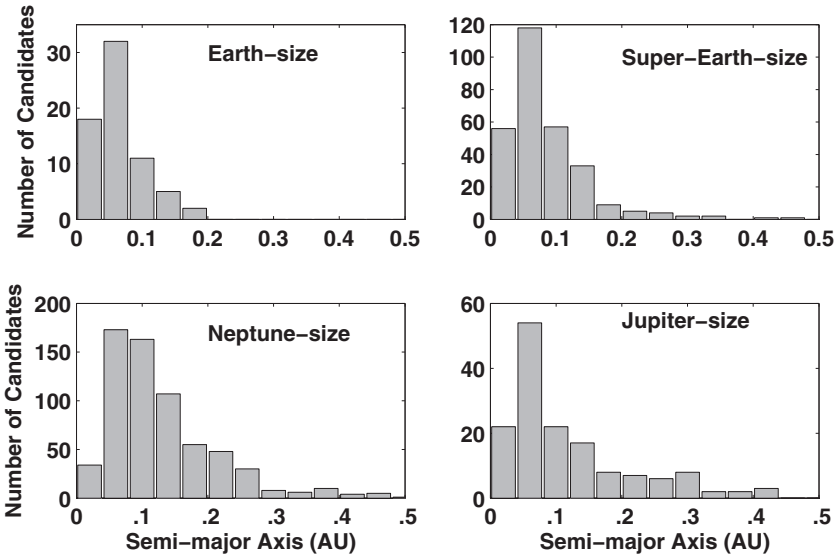

Figure 7. Number of observed candidates vs. semimajor axis for four candidate size ranges. As defined in Table 6, Earth-size refers to $R_p < 1.25\,R_\oplus$, super-Earth-size to $1.25\,R_\oplus < R_p < 2\,R_\oplus$, Neptune-size to $2\,R_\oplus$, $< R_p < 6\,R_\oplus$, and Jupiter-size refers to $6\,R_\oplus < R_p < 15\,R_\oplus$. Bin size for the semimajor axis is 0.04 AU.

Table 6
Number of Candidates vs. Size

Candidate Label	Candidate Size ($R_\oplus$)	Number of Candidates Plus Known Planets
Earth-size	$R_p \leqslant 1.25$	68
Super-Earth-size	$1.25 < R_p \leqslant 2.0$	288
Neptune-size	$2.0 < R_p \leqslant 6.0$	662
Jupiter-size	$6.0 < R_p \leqslant 15$	165
Very large size	$15.0 < R_p \leqslant 22.4$	19
Not considered	$R_p > 22.4$	15

the ease of detecting such large candidates, the difference in the dependence of number on semimajor axis is likely to be real. All show maxima in the number of candidates for orbital periods between 2 and 5 days and a narrow dip at periods shorter than 2 days. This dip is not seen for the very large candidates. However, these objects are as large as late M-dwarf stars and it is unclear what type of object they represent. Determination of their masses with RV techniques is clearly warranted because the results would not only provide masses, but densities as well when combined with the transit results.

A breakout of the number of candidates versus semimajor axis is shown in Figure 7 using the definition for size in Table 6. "Earth-size" candidates and some of the "super-Earth-size" candidates are expected to be rocky-type planets without a hydrogen–helium atmosphere. "Neptune-size" candidates could be similar to Neptune and the ice giants in composition. All size-classes show a rise in the number of candidates for decreasing semimajor axis until a value of 0.04 AU and then a steep drop. The drop off in the number of Earth-size candidates for semimajor axes greater than 0.2 AU is due at least in part to the decreasing probability of a favorable geometrical alignment and the difficulty of detecting small planets when only a few transits are available.

Figure 8 compares the orbital period distribution of the *Kepler* planet candidates with the planets discovered by the RV method (as reported by the EPE). Both detection methods show a prominent peak in the numbers for periods between 2 and 4 days and a large drop in the number for shorter periods. There are several references in the literature to the pile-up of giant planet orbital periods near 3 days (e.g., Santos & Mayor 2003). It is suggestive of a process that allows planets migrating inward to synchronize their orbital period with the rotation period of the star, raise tides of sufficient strength that enough momentum is transferred to the planet to halt its migration. Later, the star becomes sufficiently luminous that the dust and gas of the accretion disk are expelled leaving the planet in a stable, but short-period orbit. The cause of the much larger relative decrease seen in the RV-discovered planets compared to that seen in the *Kepler* results is not understood.

The planetary candidates observed at shorter distances could represent those that did not come into synchronism with the star, but stopped short of entering the star's atmosphere because of a coincidence with the dissipation of the accretion disk. They could also represent a continued migration of the body into the star.

Except for the peak between 2 and $4\,R_\oplus$, Figure 9 shows that the number of short-period (<3 days) candidates is nearly independent of candidate size through $16\,R_\oplus$. However, small candidates are more numerous than large ones for longer orbital periods. This distribution suggests that short-period candidates might represent a different population than the populations at larger orbital periods and semimajor axes. In particular, they might represent rocky planets and the remnant cores of ice giants and gas giant planets that have lost their atmospheres. To confirm that this population is distinct from that of longer-period candidates will require a future investigation of the comparison of the mass–radius relationships of the populations.

In Figure 10, the observed frequency of candidates in each magnitude bin has been simply calculated from the number of candidates in each bin divided by the total number of stars

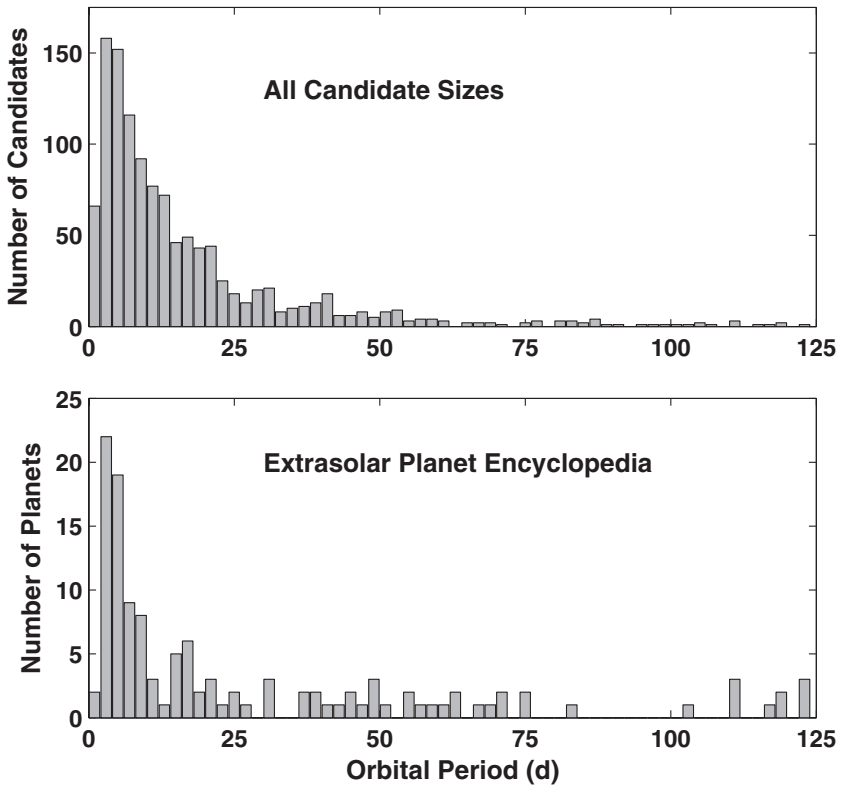

Figure 8. Upper panel: observed period distribution of *Kepler* planet candidates with orbital periods less than 125 days, uncorrected for observational selection effects. Lower panel: period distribution over the same range for RV-discovered planets listed in the Extrasolar Planets Encyclopedia (EPE) as of 2010 December 7 exclusive of *Kepler* planets. Bin size is 2 days.

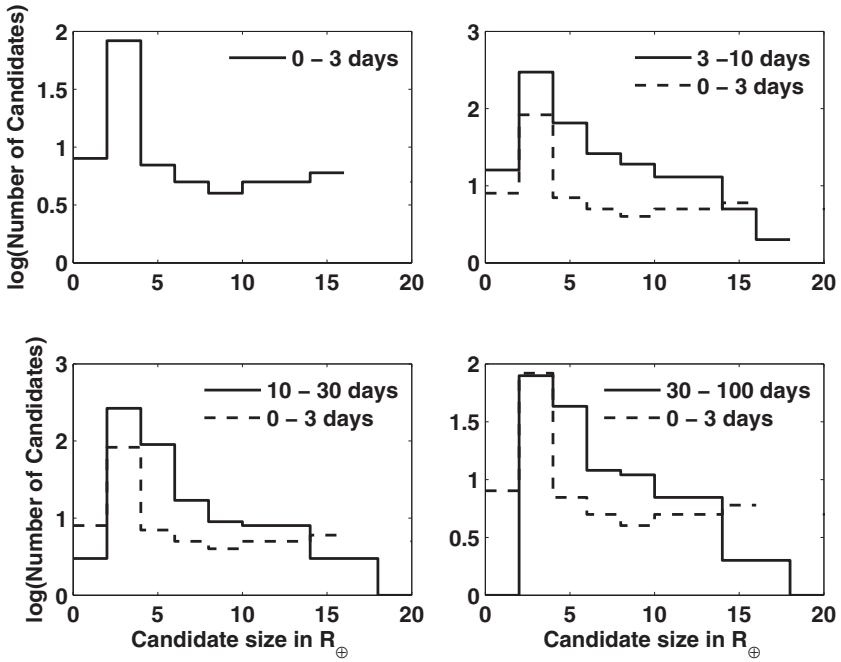

Figure 9. Observed distribution of candidate sizes for four ranges of orbital period, uncorrected for selection effects. Panels (2)–(4) compare the distributions for longer periods with that of the shortest period range. Bin size is $2\,R_{\oplus}$.

monitored in each bin. The number of stars brighter than $K_{\rm p} = 9.0$ or fainter than $K_{\rm p} = 16.0$ in the current list is so small that the count is not shown.

The panels for Earth-size and super-Earth-size candidates are consistent with a decrease in the observed frequency with increasing magnitude for magnitudes larger than $K_{\rm p} = 11$ and are indicative of difficulty in detecting small candidates around faint stars. Near-constant values of observed frequencies of the Neptune-size and larger candidates would be expected if the survey were mostly complete for the large candidates and for

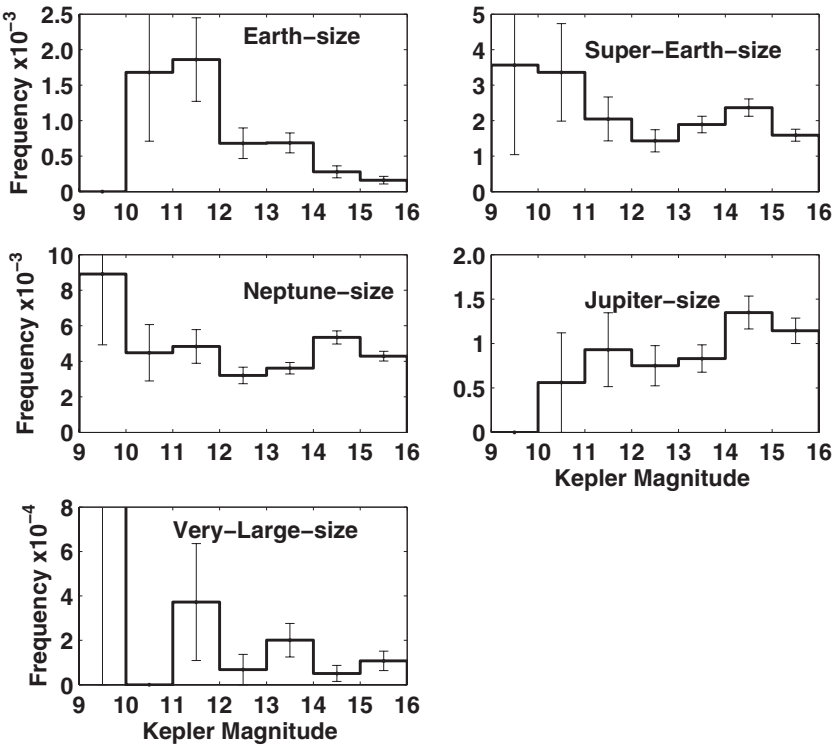

Figure 10. Observed frequencies, uncorrected for selection effects, of candidates for five size ranges defined in Table 6 as a function of *Kepler* magnitude. The error bars represent only the Poisson noise associated with the number of events in each bin, and the upper bar represents a single event if no events are observed.

the orbital periods reported here and if the distribution of stellar types is independent of apparent magnitude. However, almost all M-dwarf stars in the *Kepler* FOV have $K_p > 14$. Therefore, if the frequency of large candidates around M-dwarfs is different than for other spectral types, then near-constant frequencies of Neptune- and larger-size candidates should not be expected. Perhaps the apparent decrease with increasing magnitude is due to this cause.

An examination of the upper left panel of Figure 10 indicates that several Earth-size candidates must be present in the 15th to 16th magnitude bin. The noise properties of the instrument are such that only the smallest stars or small stars with short-period candidates can appear in this bin. To get a measure of the variation of the observed frequency distributions with magnitude when the transit amplitude is held nearly constant, the distributions for five ranges of the ratio R_p/R^* are displayed in Figure 11.

The five ratios shown in Figure 11 are appropriate for Earth-size, super-Earth-size, Neptune-size, Jupiter-size, and very large size candidate transiting stars of radius $R^* = 1\ R_\odot$, where the subscript "$\odot$" signifies solar values. An examination of the upper left-hand panel shows that no candidates are found for magnitudes between 15 and 16. The Earth-size candidates around faint stars ($K_p > 15$) shown in the upper left panel of Figure 10 orbit small stars and have a planet–star radius ratio greater than 0.0115. Thus, they no longer appear in the upper left panel of Figure 11. The observed frequency distributions show a steeper decrease with increasing magnitude for the small R_p/R^* shown in the two upper panels. The panels in the second row again show a nearly constant frequency with magnitude implying that such signal levels are readily detected over the magnitude range of interest. Contrary to what might be expected, a nearly constant frequency with magnitude is not seen for the largest ratio range. This result is not understood.

The number of candidates is a maximum for stars with temperatures between 5000 and 6000 K, i.e., G-type dwarfs (Figure 12). This result should be expected because the selection process explicitly emphasized these stars and because G-type stars are a large component of magnitude-limited surveys of dwarfs at the magnitudes of interest to the *Kepler* mission.

To reduce the bias associated with the large fraction of K-, G-, and F-type stars, the number of candidates in each bin was normalized to the number of star in the bin and frequencies calculated as a function of stellar temperature. However, because of the narrow-width temperature bins, many of the bins have a very small number of candidates which cause the frequencies to vary widely due to small-number statistics. To increase the number in each bin and reduce the large variations associated with small-number statistics, the bins in Figure 13 are twice as large as those in Figure 12.

In Figure 13, a comparison of the frequencies of super-Earth-size and Neptune-size candidates shows an indication that candidates are preferentially found around stars cooler than 4000 K. A similar distribution is also found for Earth-size candidates, but because of the very small number of candidates in that bin (i.e., 2), the maximum is not statistically significant. Main-sequence stars with temperatures between 3000 and 4000 K are classified as M-dwarfs. Giant and supergiant late K spectral-type stars are both more massive and larger than the M-dwarfs but have similar temperatures. A check of the KIC showed that none of the candidates were associated with log g less than 4.2, i.e., they are associated with dwarfs, not giants. Because M-dwarfs are much smaller than earlier spectral types, the amplitudes of the transits generated by small planets are substantially larger than those generated by hotter stars. This fact introduces a strong bias that will be considered in the next section.

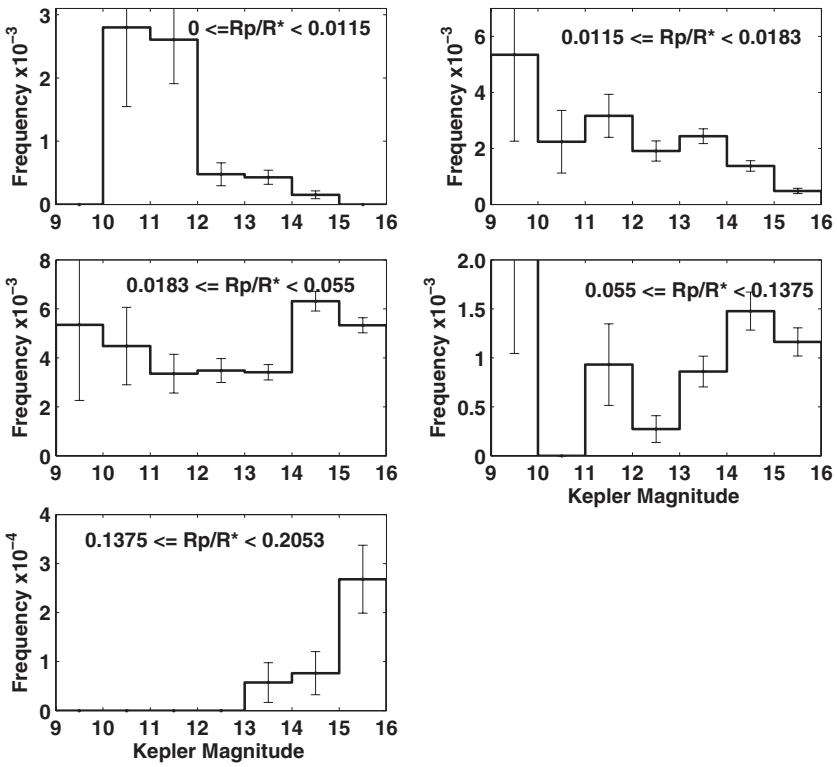

Figure 11. Frequency distribution (not corrected for selection effects) for five ranges of the ratio of the radius of the candidate to that of the host star vs. magnitude.

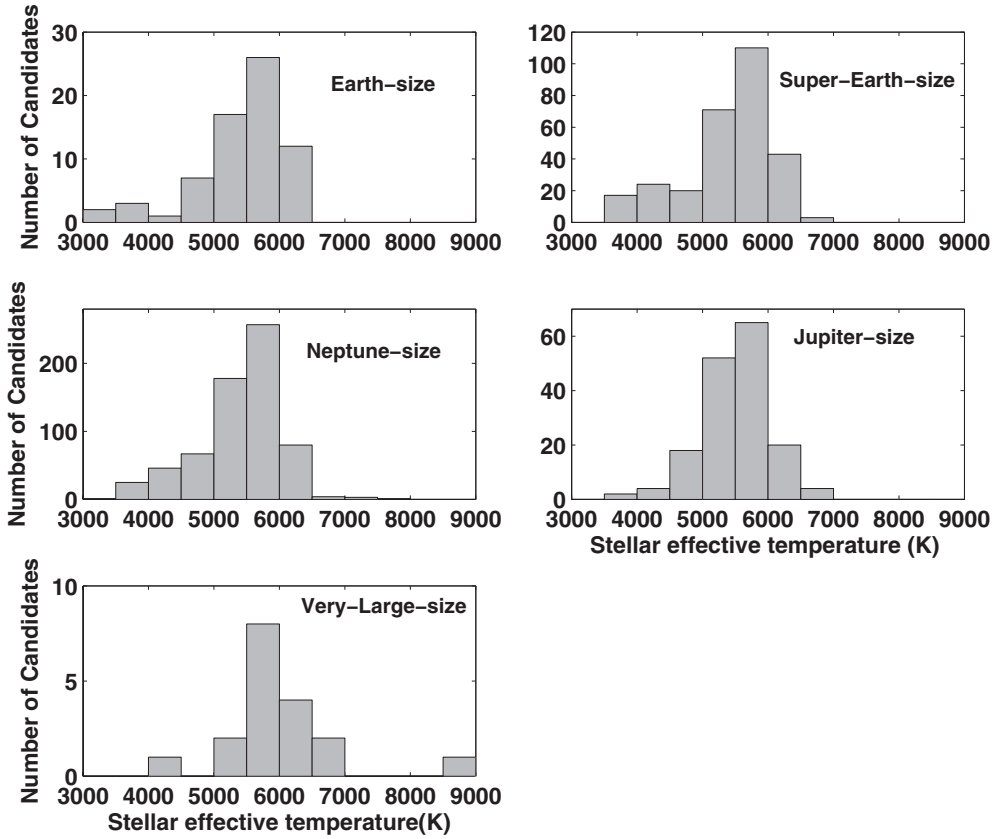

Figure 12. Observed number of candidates for various candidate sizes vs. stellar effective temperature, uncorrected for selection effects. Bin size is 500 K. Refer to Table 6 for the definition of each size category.

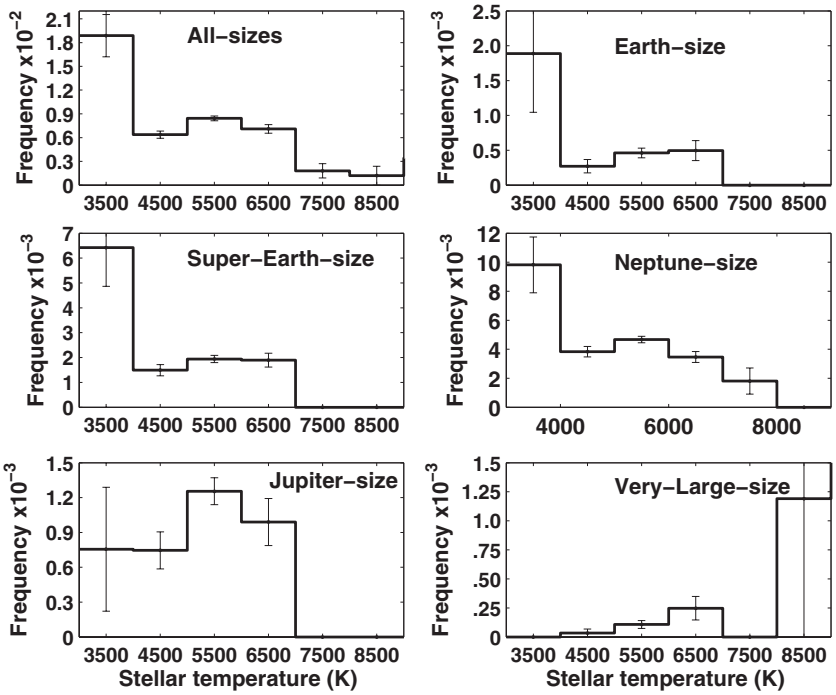

Figure 13. Measured frequency of candidates vs. stellar temperature. The error bars shown with the distributions represent only that portion of the uncertainty due to Poisson noise. Bin size is 1000 K. Refer to Table 6 for the definition of each size category.

4. COMPLETENESS ESTIMATE

Although the primary purpose of the paper is to summarize the results of the observations and to act as a guide to content of the tables, a model was developed to provide a first estimate of the intrinsic frequency of planetary candidates. The "intrinsic frequency" of planetary candidates is used here to mean the observed number of candidates per number of target stars that must be observed to produce the observed number of candidates in the specified bins of semimajor axis a and candidate size R when all selection effects are applied. The bin limits used for a are evenly spaced from 0.0 to 0.5 AU with a spacing of 0.02 AU. The bin limits for the planetary candidate size-classes are Earth-size ($0.5\,R_\oplus \leqslant R < 1.25\,R_\oplus$), super-Earth-size ($1.25\,R_\oplus \leqslant R < 2.0\,R_\oplus$), Neptune-size ($2.0\,R_\oplus \leqslant R < 6.0\,R_\oplus$), Jupiter-size ($6.0\,R_\oplus \leqslant R < 15.0\,R_\oplus$), and very large size ($15.0\,R_\oplus \leqslant R < 22.4\,R_\oplus$). It should be noted that the calculation of the intrinsic frequency is equivalent to the ratio of the measured number of candidates divided by the expected number of candidates based on the ensemble of stars that are observed.

For every candidate in a $\Delta a\ \Delta R$ bin, each of the 153,196 target stars was examined to determine if a planet orbiting it with the same size as the candidate and having the same a could be detected during the Q0 through Q2 observation period. The number of target stars needed to produce a minimum of two transits in the period of interest with a signal $\geqslant 7\sigma$ was tabulated for each bin. (There is no need for three transits because confirmation as a planet is not considered here.) The actual period simulated is longer than the 138 days of the Q0 through Q2 period because the search for planetary candidates used data obtained during later periods to obtain accurate values of the epoch and period, as discussed earlier.

Inputs to the model include the observed noise for 3, 6, and 12 hr bins averaged over one quarter of data (Q3) for each target star and the target star's size, mass, and magnitude, as well as the values of the size and semimajor axis of each candidate in the $\Delta a\ \Delta R$ bin. We also undertook an independent analysis that used the observed noise for 3 hr bins averaged over the Q3 data. Since the properties of the noises are not Gaussian, this serves as a check on our results.

The model computes the duration of the transits from the size and mass of the star at the specified value of the semimajor axis. The value of the noise for each target star is interpolated to the computed transit duration based on the values of the noise measured for 3, 6, and 12 hr samples. This a very important correction because for 80% of the stars, the variation of CDPP with the duration of the transit does not vary with the reciprocal of the square root of the time, but is less than that expected from a Poisson distribution. The signal level is computed from the square of the ratio of the candidate size to the size of the target star. This value is then divided by the interpolated noise value to get the estimated single-transit S/N. The total S/N is based on the single-transit S/N multiplied by the square root of number of transits that occur during the observation period. A correction is made for the loss of transits (and consequently, the reduction in the total S/N) due to the monthly and quarterly interruptions of observations. The probability of a recognized detection event is then computed from the value of the total S/N and a threshold level of 7σ. In particular, if the total S/N is 7.0, then the transit pattern will be recognized 50% of the time while if the total S/N was estimated to be 8.0, then the transit pattern would be recognized 84% of the time. The value of this probability p_1 is tabulated and then an adjustment is made for the probability that the planet's orbit is correctly aligned to the line-of-sight p_2. The value of p_2 is based on the size of the target star and the semimajor axis specified for the candidate. The product of these probabilities $p_{\rm nc}$ is the probability that the target star n could have produced the observed candidate c.

The probability $p_{\rm nc}$ is computed for each of the 153,196 stars and then summed to yield the estimated number of target stars $n^*_{c,a,R}$ that could have produced a detectable signal consistent with candidate's semimajor axis a and size R. (Subscripts

designate candidate "c," semimajor axis value "a," candidate size "R.") This procedure is repeated for each candidate in the $\Delta a\ \Delta R$ bin.

The sum of the number of candidates of size-class "k" in a bin (a, Δa, R, ΔR) is designated $S_{a,R,k}$. The size-class "k" ($k = 1$–5) represents Earth-size, super-Earth-size, Neptune-size, Jupiter-size, and very large size planetary candidates, respectively.

After a value of $n^*_{c,a,R,k}$ has been computed for each candidate in the bin, the median value $N^*_{a,R,k}$ of $n^*_{c,a,R,k}$ is computed and used to estimate the frequencies:

$$\mathrm{Freq}(k, a_i, \Delta a_i, R_i, \Delta R_i) = \frac{S_{a,R,k}}{N^*_{a,R,k}}. \quad (1)$$

For each size-class, the sum of the frequencies over a and R is the estimate of the frequency for that size-class:

$$\mathrm{Freq}(k) = \sum_{R=\min}^{R=\max} \sum_{a=0.01}^{a=0.5} \frac{S_{a,R,k}}{N^*_{a,R,k}}. \quad (2)$$

The summation for each size-class is done only for those bins that have at least 2 planetary candidates and a minimum of 10 target stars. These choices help to reduce the impact of outlier values.

The uncertainties in the results are quite large because the calculated number of stars $n^*_{c,a,R,k}$ for the observed number of candidates $S_{a,R,k}$ is a sensitive function of the position of each planetary candidate inside of the $\Delta a\ \Delta Rr$ bin and because the number of candidates in each bin is often small. In particular, estimated frequencies based on the sum of the individual frequencies in each bin are very different than the estimates obtained by dividing the number of observed candidates by the average number of expected planets. Therefore, medians are used instead of averages to reduce the effects of outliers.

To provide an estimate of the dispersion $D_{a,R,k}$ of the estimated frequencies for each bin, the relative error associated with the number of candidates used in the estimate of the frequency is added in quadrature to the variance due to the dispersion of the values of $n^*_{c,a,R,k}$:

$$D_{a,R,k} = \sqrt{\frac{1}{S_{a,R,k}} + \frac{\mathrm{Var}(n^*_{c,a,R,k})}{\overline{(n^*_{c,a,R,k})^2}}}, \quad (3)$$

$$\text{where} \quad n^*_{c,a,R,k} = \sum_{c=1}^{c=\max} n^*_{c,a,R,k}. \quad (4)$$

It is important to note that the estimated frequencies calculated by the model are based upon the number of candidates found in the data. In turn, the number and size distributions depend on both the results from the analysis pipeline and a manual inspection of the results of the pipeline product. The current version of the analysis pipeline provides "threshold crossing events" and checks that those data are consistent with an astrophysical process. However, it does not yet have the capability to stitch together quarterly records. Thus, the number of candidates discussed here is based on a combination of pipeline results, manual inspection, and an ad hoc program that does not use the more comprehensive detrending that is done in the pipeline, but does allow a longer period of data to be examined. In some cases, the candidates in the Q0–Q2 data were not discovered until the Q3 and Q5 data were examined. As discussed later, the procedure is designed to quickly find candidates that can be followed up, but is not well controlled for the purpose of the model calculations. Consequently, the results must be considered very preliminary.

Table 7 presents an example of the calculated intrinsic frequencies, number of planetary candidates, mean value of the number of target stars, and dispersion values for the range of a from 0.01 to 0.50 AU for Earth-size candidates. The results for the all class-sizes are plotted in Figure 14.

The estimated intrinsic frequencies summed over semimajor axis are 0.05, 0.08, 0.18, 0.02, and 0.001 for Earth-, super-Earth-, Neptune-, Jupiter- and very large size planetary candidates, respectively. The sum over all values of the semimajor axis is 0.34. This value is interpreted to mean that the average number of candidates per star with semimajor axes less than 0.5 AU is 0.341 with a very large uncertainty.

When the model is run to simulate a six-month period, the results are very similar for candidates Neptune-size and larger, but the frequencies of super-Earth and Earth-size candidates are increased by 3 for Earth-size candidates and 2 for super-Earth-size candidates. The uncertainty in the predictions will decrease as the mission duration increases and the number of transits and resulting S/N increase.

All the panels in Figure 14 show a large increase in intrinsic frequency with semimajor axis from the 0.00 to approximately 0.07 AU and then show a negative or near-zero slope at larger values of the semimajor axis. (The variation of intrinsic frequency for the very large candidates is too noisy to characterize.) The result for the Jupiter-size candidates shows a nearly constant value with semimajor axis. The peak in the intrinsic frequencies for the three smallest class-sizes is located in the bin to the immediate right of the peak in the observations. The distribution of the "All Sizes" class is distorted by the lack of the values of Earth- and super-Earth-sizes for semimajor axes greater than 0.15 AU and 0.25 AU, respectively.

In Figure 15, the dependence of the intrinsic frequencies on the stellar temperature is examined. Note that these results subsume the entire range of semimajor axis just discussed.

The results shown in Figure 15 indicate that once adjustments are made for the increased sensitivity to small planets orbiting small stars as opposed to Sun-like stars, the higher frequency of Earth-size candidates orbiting the coolest stars seen in Figure 13 disappears. However, the peak for super-Earth-size and Neptune-size is still prominent and it is also clear that the Jupiter-size and very large candidates are much more frequent around hotter than they are for the cooler M- and K-type stars.

An examination of the panel in Figure 15 for the frequency dependence of Neptune-sized candidates suggests a negative correlation with temperature. The linear correlation coefficient has a value of -0.95 with 95% confidence limits for the coefficient between -0.995 and -0.57. Although the intrinsic frequencies of Jupiter-sized and very large sized candidates also suggest a correlation with stellar effective temperature, because of the small number of data points, no formal estimation can be obtained for their correlation coefficients nor those for the Earth-size and super-Earth-size candidates.

One of the surprising results shown in Figure 15 is the dip in the intrinsic frequency of Earth-size and super-Earth-size candidates orbiting stars with temperatures near 4500 K, i.e., K-type stars. A careful inspection of the lower-left panel of Figure 3 also shows a paucity of candidates for temperatures between 4000 and 5000 K. The large values of the dispersion

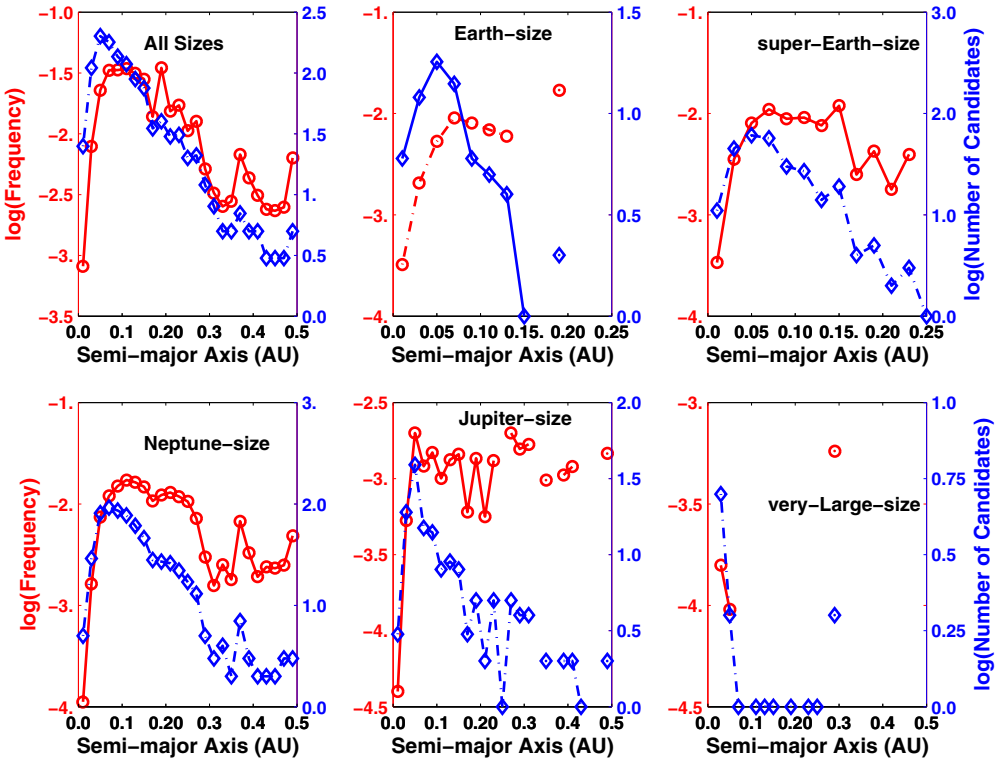

Figure 14. Comparisons of the logarithms of intrinsic frequencies "log(frequency)" to observations "log(No. of candidates)" as a function of semimajor axis for five size classes. Red symbols (circles) denote intrinsic frequencies and use the scales on the left vertical axes. Blue symbols (diamonds) denote the number of candidates and use the scales on the right vertical axes. To reduce the effect of outliers, values for the intrinsic frequencies are shown only when at least two candidates are found in the bin. Frequencies are based on 0.02 AU bins.

Table 7
Intrinsic Frequency of Earth-size Candidates (Simulation of 1.0 Year of Observations)

Results for Earth-size Candidates					
a (AU)		$S_{a,R,k=1}$	$N_{a,R,k=1}$	Frequency	Relative Dispersion
0.001	0.02	6	18551.7	3.23E-04	0.62
0.02	0.04	12	5815.9	2.06E-03	0.55
0.04	0.06	18	3400.4	5.29E-03	0.5
0.06	0.08	14	1541.1	9.08E-03	0.7
0.08	0.1	6	744.9	8.05E-03	0.65
0.1	0.12	5	722	6.92E-03	0.68
0.12	0.14	4	667.1	6.00E-03	0.55
0.14	0.16	1	0	0.00E+00	1
0.16	0.18	0	0	0.00E+00	0
0.18	0.2	2	117.7	1.70E-02	0.79
0.2	0.22	0	0	0.00E+00	0
0.22	0.24	0	0	0.00E+00	0
0.24	0.26	0	0	0.00E+00	0
0.26	0.28	0	0	0.00E+00	0
0.28	0.3	0	0	0.00E+00	0
0.3	0.32	0	0	0.00E+00	0
0.32	0.34	0	0	0.00E+00	0
0.34	0.36	0	0	0.00E+00	0
0.36	0.38	0	0	0.00E+00	0
0.38	0.4	0	0	0.00E+00	0
0.4	0.42	0	0	0.00E+00	0
0.42	0.44	0	0	0.00E+00	0
0.44	0.46	0	0	0.00E+00	0
0.46	0.48	0	0	0.00E+00	0
0.48	0.5	0	0	0.00E+00	0
0.5	0.52	0	0	0.00E+00	0

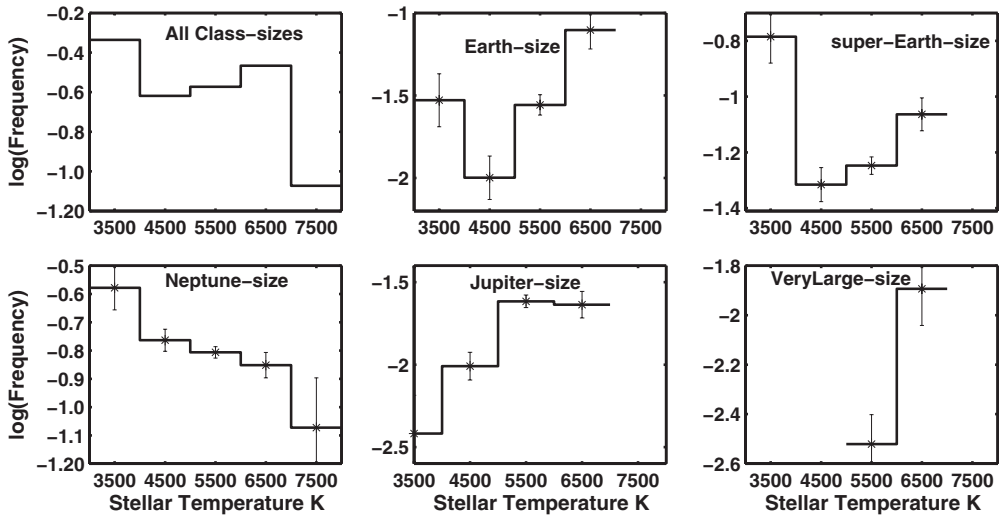

Figure 15. Logarithm of the intrinsic frequencies as a function of stellar effective temperature after implementing the sensitivity corrections described in Section 4. The bins along the *x*-axis span 3000–4000 K, 4000–5000 K, 5000–6000 K, and 6000–7000 K, with each bin labeled by the central value.

shown in Figure 15 indicate that the result should be interpreted with caution.

It should be noted that the values for the intrinsic frequencies in Table 7 and in Figures 14 and 15 must be considered preliminary estimates. These values will be lowered when more false positive events are recognized and removed, but they could also increase; the precision of the data is assumed to improve as the square root of the number of measurements in transit. If, however, the performance of the data does not achieve this ideal case, then fewer stars are being searched than assumed here. Thus, the inherent frequency would be higher than shown in Table 7 and associated figures. Furthermore, throughout the mission we will continue to make improvements to the data analysis pipeline. As the capability of the system to recognize small candidates improves, and more candidates in the data discussed here will be discovered. A significant improvement is expected in mid-year when the capability to stitch together quarters of observations becomes operational.

It is interesting to compare these results with those of Howard et al. (2010) for planets with periods $\leqslant$50 days discovered by RV. For planet masses 3–10 $M_{\oplus}$ (super-Earth-mass), they get approximately 10.7%–11.8% while the present calculation for candidates with comparable periods and for super-Earth-size gives 8%. For 10–30 $M_{\oplus}$, Howard et al. obtain 5.8%–6.5% while the *Kepler* results for Neptune-size candidates predict 18%. The agreement is satisfactory given the many uncertainties involved in the estimates and the fact that size estimates are being compared to mass estimates.

5. OVERVIEW OF MULTI-PLANET SYSTEMS

A total 170 target stars with multiple planet candidates have been detected among the 997 host stars in *Kepler* data. There are 115 stars with two candidates, 45 with three candidates, 8 stars with four candidates, 1 star with five, and 1 star with six candidates. For these figures all candidates are included, whether they are validated planets or not. The fraction of host stars that have multi-candidate systems is 0.17 and the fraction of the candidates that are part of multi-candidate systems is 0.339, i.e., 408 among 1202 candidates. Because all the candidates discussed here show two or more transits, accurate orbital periods and epochs are available in Table 2.

Comparisons of the distributions presented in Figure 16 with previous figures show that they are similar to those for the ensemble of all candidates. The number versus orbital period is very much like that seen in Figure 6: a lack of candidates with orbital periods less than 2 days, a maxima near 4 days, and a gradual reduction in the number with orbital period. The number versus candidate size in Figure 16 is quite similar to that in Figure 2. The peak in the frequency with stellar temperature for cool stars is also repeated. However, the distributions displayed in the two scatter plots in the middle panel of Figure 16 show that the size versus orbital period and semimajor axis are different from those in Figure 3. In particular, both of the distributions shown in Figure 16 display a lack of giant planets for close-in/short-period orbits compared to the distributions in Figure 3. There is a clear paucity of giant planets in the observed multi-candidate and multi-planet systems (see Latham et al. 2011 for details). This result is consistent with RV surveys which indicate that short-period giant planets are significantly less common in multiple planet systems (Wright et al. 2009).

An unusual candidate, KOI 961.02, is shown in the second row, left-hand panel of Figure 16. It has a period of 0.45 days, a semimajor axis of 0.01 AU, and a size 28% larger than Jupiter. So far it has passed all vetting tests and will be on the list to get an RV confirmation.

Multiple planet candidate systems, as well as the single-planet candidate systems, could harbor additional planets that do not transit or have not yet been recognized as such, and therefore are not seen in these data. Such planets might be detectable via TTVs of the transiting planets after several years of *Kepler* photometry (Agol et al. 2005; Holman & Murray 2005; Holman et al. 2010). A preliminary analysis of transit times of planetary candidates based on data up to and including quarter 2 provides hints that ∼65 KOIs may already exhibit TTVs. A statistical analysis of these and many other marginal TTV signals has been submitted (Ford et al. 2011). Papers with TTV confirmation of three systems are already published (Holman et al. 2010; Lissauer et al. 2011a) or in preparation (Cochran et al. 2011). Ford et al. (2011) predicts that *Kepler* will confirm (or reject) at least ∼12 systems with multiple transiting planet candidates via TTVs.

It is important to note that it is possible, though unlikely, for light from more than one background EB star system to be

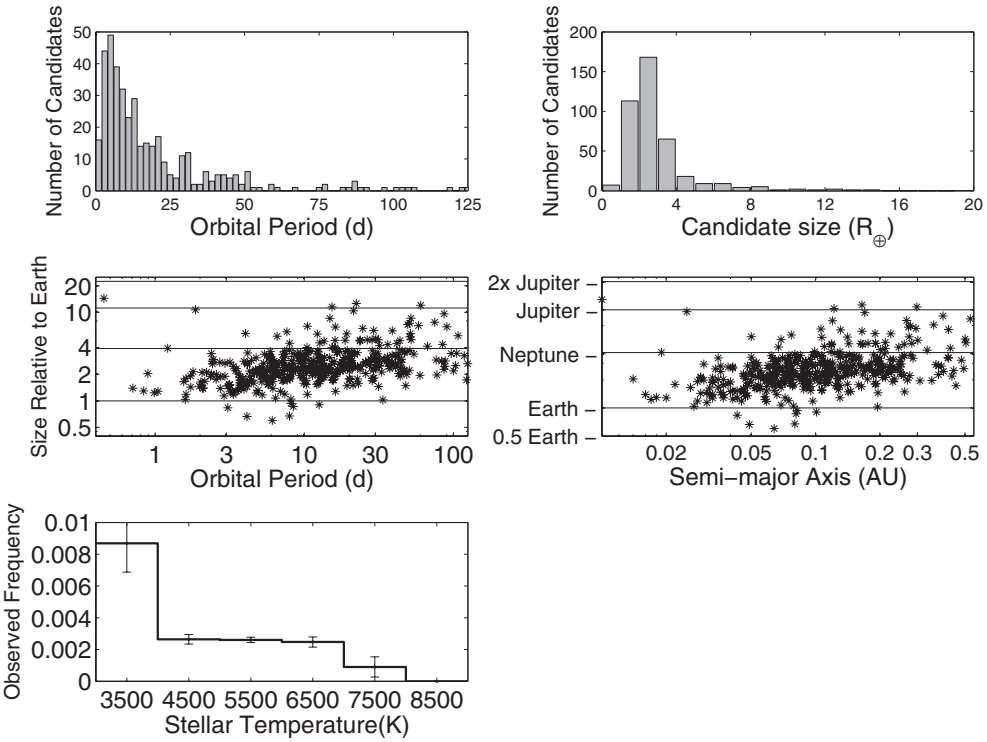

Figure 16. Observed distributions of planetary candidates in multi-planet candidate systems. Bin sizes for the upper two panels and the lower panel are 2 days, 1 $R_\oplus$, and 1000 K, respectively. Refer to Table 6 for the definition of each size category.

within the photometric aperture, producing an apparent multi-planet transit signal in the light curve. While Latham et al. (2011) and Lissauer et al. (2011b) present several arguments showing that candidates in multiples are more likely to be true planets, a thorough analysis of each system and a check of background binaries are required before any discovery can be claimed. Approximately 34% of *Kepler* candidates are part of multi-candidate systems. The corresponding fraction of RV planets in multi-planet systems is 30% based on the EPE. The fraction of stars with multiple known planets or candidates is 17% for the *Kepler* sample and about 12% for the RV sample. Given the various limitations of these two observing techniques, these numbers are consistent. While an exhaustive study remains to be done, Lissauer et al. (2011b) investigated the dynamical attributes of *Kepler* multi-candidate systems and also suggest that nearly coplanar planetary systems might be common.

6. SUMMARY AND CONCLUSIONS

Distributions of the characteristics of 1202 planetary candidates have been given. These include number and frequency distributions with orbital size and period, stellar temperature, and magnitude. These distributions are separated into five class-sizes: 68 candidates of approximately Earth-size ($R_p < 1.25\,R_\oplus$), 288 super-Earth-size ($1.25\,R_\oplus \leqslant R_p < 2\,R_\oplus$), 662 Neptune-size ($2\,R_\oplus \leqslant R_p < 6\,R_\oplus$), 165 Jupiter-size ($6\,R_\oplus \leqslant R_p < 15\,R_\oplus$), and 19 up to twice the size of Jupiter ($15\,R_\oplus \leqslant R_p < 22\,R_\oplus$). Over the temperature range appropriate for the HZ, 54 candidates are found with sizes ranging from Earth-size to larger than that of Jupiter. Six planetary candidates in the HZ are less than twice the size of the Earth.

Over 74% of the planetary candidates are smaller than Neptune. The observed number versus size distribution of planetary candidates increases to a peak at two to three times the Earth-size and then declines inversely proportional to the area of the candidate. For candidate sizes greater than $2\,R_\oplus$, the dependence of the number of candidates on the candidate radius is proportional to the reciprocal of the square of the candidate radius.

However, there is a prominent decrease in the number of candidates with size in all class-sizes for semimajor axes smaller than 0.07 AU and for orbital periods less than 3 days. A group of candidates with orbital periods less than 3 days is identified that appears distinctly different from those with longer periods. In particular the size distribution of candidates with short orbital periods is nearly constant with candidate size.

The intrinsic frequencies of super-Earth-size and Neptune-size candidates show maxima for the coolest stars. Both Earth-size and super-Earth-size candidates show minima for stars with temperatures near 4500 K. Jupiter-size and very large size candidates show much higher frequencies for hotter stars than for those cooler than 5500 K.

The analysis of the first four months of *Kepler* observations is the first to estimate the frequency of small candidates (Earth-size, super-Earth-size, and Neptune-size) based on a uniform set of observations with the capability of detecting small candidates. After correcting for geometric and sensitivity biases, we find intrinsic frequencies of 5% for Earth-size candidates, 8% for super-Earth-size candidates, 18% for Neptune-size candidates, and 2% for Jupiter-size candidates.

Multi-candidate, transiting systems are frequent; 17% of the host stars have multi-candidate systems, and 34% of all the candidates are part of multi-candidate systems.

There is also evidence for 34 candidates with sizes between 1.3 and 4.5 times that of Jupiter. The nature of these candidates is unclear. The 19 that are between 1.3 and 2.0 times the size of Jupiter are included in tables and figures presented in this paper because of the possibility that they are very inflated planetary objects, but the 15 larger than twice the size of Jupiter were omitted from the discussion because it is more likely that they

are stellar objects or that the estimated size of the host star is much smaller than listed in the KIC.

In the coming years, many of these candidates are expected to be reclassified as exoplanets as the validation effort proceeds. The number of candidates is so large that the *Kepler* team must be selective in its follow-up program and will devote the majority of its efforts to the detection and validation of the smallest candidates and to those with orbital periods appropriate for the HZ and those amenable to follow-up. Many candidates will be left to future work or for follow-up by the community. The release of the Q0 through Q1 data and the early release of the Q2 data and the descriptions of the candidates with accurate positions, magnitudes, epochs, and periods should help the community to confirm and validate many of these candidates.

The data released here should also provide to the community a more comprehensive source of data and distributions needed for further developments of the theories of planet structure and planetary systems. These results have concentrated upon discovery of candidates, and initial levels of validations sufficient to cull out many false positives. Future studies by the *Kepler* science team will include efforts to robustly quantify the completeness of these candidate lists through simulation studies and provide more refined confidence levels on probabilities of candidates being planets. Discovery of additional candidates will of course continue and should reduce the incompleteness for weak signals whether those follow from small planets, long orbital periods, or faint stars.

The *Kepler* Mission was designed to determine the frequency of extrasolar planets, the distributions of their characteristics, and their association with host star characteristics. The present results are an important milestone toward the accomplishment of *Kepler*'s goals.

Kepler was competitively selected as the tenth Discovery mission. Funding for this mission is provided by NASA's Science Mission Directorate. Some of the data presented herein were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California, and the National Aeronautics and Space Administration. The Keck Observatory was made possible by the generous financial support of the W. M. Keck Foundation. We sincerely thank Andrew Gould for his timely, thorough, and very helpful review of this paper. The authors thank many people who gave so generously of their time to make this mission a success.

REFERENCES

Agol, E., Steffen, J., Sari, R., & Clarkson, W. 2005, MNRAS, 359, 567
Bakos, G. Á., et al. 2010, ApJ, 710, 1724
Batalha, N. M., et al. 2010a, ApJ, 713, L109
Batalha, N. M., et al. 2010b, ApJ, 713, L103
Batalha, N. M, et al. 2011, ApJ, 729, 27
Borucki, W. J., et al. 2009, Science, 325, 709
Borucki, W. J., et al. 2010, ApJ, 713, L126
Borucki, W. J., et al. 2011, ApJ, 728, 117
Caldwell, D. A., et al. 2010, ApJ, 713, L92
Drimmel, R., Cabrera-Lavers, A., & Lopez-Corredoira, M. 2003, A&A, 409, 205
Dunham, E. W., et al. 2010, ApJ, 713, L136
Ford, E. B., et al. 2011, ApJ, submitted (arXiv:1102.0544)
Gautier, T. N., et al. 2010, arXiv:1001.0352
Gilliland, R. L., et al. 2010, ApJ, 713, L160
Gould, A., Pepper, J., & DePoy, D. L. 2003, ApJ, 594, 533
Heath, M. J., Doyle, L. R., Joshi, M. M., & Haberle, R. M. 1999, Orig. Life Evol. Biosph., 29, 405
Holman, M. J., & Murray, N. W. 2005, Science, 307, 1288
Holman, M. J., et al. 2010, Science, 330, 51
Howard, A. W., et al. 2010, Science, 330, 653
Jenkins, J. M. 2002, ApJ, 575, 493
Jenkins, J. M, et al. 2010a, ApJ, 724, 1108
Jenkins, J. M., et al. 2010b, ApJ, 713, L120
Jenkins, J. M., et al. 2010c, ApJ, 713, L87
Jenkins, J. M., et al. 2010d, Proc. SPIE, 7740, 77400D
Joshi, M. 2003, Astrobiology, 3, 415
Kaltenegger, L., & Sasselov, D. 2011, ApJL, in press (arXiv:1105.0861)
Kasting, J. E., Whitmire, D. P., & Reynolds, R. T. 1993, Icarus, 101, 108
Koch, D. G., et al. 2010a, ApJ, 713, L79
Koch, D. G., et al. 2010b, ApJ, 713, L131
Latham, D. W., et al. 2010, ApJ, 713, L140
Latham, D. W., et al. 2011, ApJ, 732, 24
Lissauer, J. J., et al. 2011a, Nature, 470, 53
Lissauer, J. J., et al. 2011b, ApJ, submitted (arXiv:1102.0543L)
Morton, T. D., & Johnson, J. A. 2011, in press (arXiv:1101.5630)
O'Donovan, F. T., et al. 2006, ApJ, 651, L61
Pál, A., et al. 2008, ApJ, 680, 1450
Prsa, A., et al. 2011, AJ, 141, 83
Raghavan, D., et al. 2010, ApJS, 190, 1
Rampino, M. R., & Caldeira, K. 1994, ARA&A, 32, 83
Santos, N. C., & Mayor, M. 2003, in Unsolved Universe: Challenges for the Future, ed. M. J. P. F. G. Monteiro (Dordrecht: Kluwer), 15
Segura, A., Kasting, J. F., Meadows, V., Cohen, M., Scalo, J., Crisp, D., Butler, R. A. H., & Tinetti, G. 2005, Astrobiology, 5, 706
Steffen, J. H., et al. 2010, ApJ, 725, 1226
Tarter, J. C., et al. 2007, Astrobiology, 7, 30
Torres, G., et al. 2011, ApJ, 727, 24
Wright, J. T., Upadhyay, S., Marcy, G. W., Fischer, D. A., Ford, E. B., & Johnson, J. A. 2009, ApJ, 693, 1084
Wu, H., et al. 2010, Proc. SPIE, 7740, 774019